CESifo WORKING PAPERS

10264
2023

February 2023

The Sequential Search Model: A Framework for Empirical Research

Raluca Ursu, Stephan Seiler, Elisabeth Honka

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10264*

The Sequential Search Model: A Framework for Empirical Research

Abstract

We provide a detailed overview of the empirical implementation of the sequential search model proposed by Weitzman (1979). We discuss the assumptions underlying the model, the identification of search cost and preference parameters, the necessary normalizations of utility parameters, counterfactuals that require a search model framework, and different estimation approaches. The goal of this paper is to consolidate knowledge and provide a unified treatment of various aspects of sequential search models that are relevant for empirical work.

JEL-Codes: D430, D830, L130.

Keywords: sequential search model.

Raluca Ursu
NYU Stern School of Business
USA – New York, NY 10012
rmu208@stern.nyu.edu

Stephan Seiler
Imperial College London, Business School
United Kingdom - London SW7 2AZ
stephan.a.seiler@gmail.com

Elisabeth Honka
UCLA, School of Management
USA - Los Angeles, CA 90095-148
elisabeth.honka@anderson.ucla.edu

This draft: January 26, 2023
Authors are listed in reverse-alphabetical order. We thank Giovanni Compiani, Rafael Greminger, Ilya Morozov, Matthijs Wildenbeest, Dan Yavorsky, and Luna Zhang for their comments. We are grateful to Weiheng Zhang and Konstantin Pavlenkov for their excellent research assistance. None of the authors received external funding for this paper. All errors are our own.

1 Introduction

Prior to shopping, consumers often have limited information about products and need to engage in costly search to gather information about possible options. The consumer search process has increasingly become observable to researchers, especially in online settings, in which browsing data allow the researcher to observe which products a consumer inspects before making a purchase. The increased availability of consumer search data has led researchers to estimate structural models of consumer search behavior which had originally been developed as theory models rather than templates for empirical work (e.g., Stigler 1961, McCall 1970, Weitzman 1979). While structural search models share some features with perfect information demand models, they involve the estimation of additional parameters, such as search costs, raise questions about the identification of these additional parameters, and pose unique computational challenges.

In this paper, we provide a detailed overview of the sequential search model proposed by Weitzman (1979), which has emerged as the most frequently used framework for empirical research on consumer search. To the best of our knowledge, arguments with regards to issues such as the identification of search cost and preference parameters, unobserved heterogeneity in search models, necessary normalizations of utility parameters, and the appropriate estimation approach have typically only been discussed in the specific context of a given application. Our aim in this paper is to consolidate knowledge and provide a general and unified treatment of the aspects of the sequential search models that are relevant for empirical work.

To set the stage, we first outline the general consumer search problem as a dynamic optimization problem with an arbitrary utility function in Section 2. We then describe the decision rules governing optimal consumer behavior and discuss necessary assumptions about the information environment. Having introduced the general search framework, we then describe the parameterizations of utility and search costs that are typically employed in empirical work.

In Section 3, we discuss the role of pre- and post-search error terms that are used in empirical models and show that one needs to normalize the variance of both error terms in many settings. This normalization prevents the researcher from quantifying search costs in monetary terms and makes it infeasible to conduct counterfactuals that alter search costs. We also characterize conditions under which the post-search normalization can be relaxed. These error normalizations have received some attention in recent work (Morozov et al. 2021, Yavorsky et al. 2021), but are, in our opinion, underappreciated relative to their importance for the interpretation of estimated model parameters.

We turn to identification in Section 4. We first discuss identification in a homogenous model with common search costs, and show how preferences are separately identified from search costs. Next, we consider the homogenous search model with product-specific search costs and show that preferences and product-specific search cost parameters can be separately identified in this model as well. Hence, both preferences and search costs can be modeled as a function of the same set of characteristics. Lastly, we discuss identification in a search model with unobserved heterogeneity in preferences based on the insight that search models provide information on the distribution of preference heterogeneity that is akin to second-choice data, which is well-known to help with the estimation of preference heterogeneity parameters (see, e.g., Berry et al. 2004).

In Section 5, we derive expressions for own- and cross-price elasticities and show that they resemble elasticities derived from full information discrete choice models, with the exception of an additional term that depends on search costs and leads to a lower price sensitivity of demand as search costs increase. Similarly, we derive a welfare expression that closely mimics the equivalent expression in the perfect information case. Finally, we discuss a series of counterfactuals that require a search model framework, such as lowering or removing search costs or altering variables that influence search costs (e.g., product rankings).

In Section 6, we discuss various methods to estimate search models such as crude and kernel-smoothed frequency estimators, an approach based on the GHK simulation method and an estimator based on importance sampling. We provide a unified notation throughout all estimation approaches to highlight their similarities and differences. We also report results from a set of Monte Carlo simulations that compare different estimation methods in terms of accuracy and computational speed. These simulations are based on accompanying codes that are publicly available.[1]

Parts of this overview are based on specific empirical search papers and we highlight the papers that inspired particular aspects of our analysis clearly throughout. We also note that certain aspects of the search literature are outside of the scope of this paper. We do not discuss the distinction between simultaneous and sequential search (De Los Santos et al. 2012, Honka and Chintagunta 2017). We confine ourselves to settings with individual-level data on search spells and purchases. We do not consider model extensions where the "one-period ahead" search rule of the canonical sequential search model may not be optimal. In addition, we do not cover learning or state dependence in search. We also do not aim to provide an exhaustive overview of the empirical search literature and only cite papers selectively when relevant in the context of a specific methodological issue. We refer the interested reader to Honka et al. (2019) for a more general overview of the search literature.

2 Consumer Search and the Weitzman (1979) Framework

In this section, we lay the foundation for our discussion of empirical work using the sequential search model. We first outline a more general version of the consumer search problem than the model specifications that have typically been used as a basis for empirical work. We then discuss the optimal decision rules derived by Weitzman (1979) and the assumptions under which the general model can be represented by Weitzman's set of optimal decision rules. Finally, we discuss parametrizations of utility and search costs that are typically used in empirical work.

2.1 General Model of Consumer Search

At a general level, consumer search is a dynamic stopping problem in which a decision maker sequentially samples options and optimally decides when to stop the search process. This stopping

[1]Link: https://drive.google.com/file/d/1-GUVl-FtF_nkKQx-A6wjG0jQNwt3fqjm/view?usp=share_link.

problem can be characterized as follows: assume a decision maker i faces a set of $\Im = \{1, \ldots, J\}$ boxes. Each box j contains a potential reward u_{ij} ("utility") independently drawn from a known distribution $F_{ij}(u)$. It costs $c_{ij} > 0$ for i to open box j and reveal its reward. The decision maker opens boxes ("searches") sequentially and her goal is to maximize her expected reward net of total costs. An outside option $j = 0$ with a known reward u_{i0} is available at no cost. The decision maker has unit demand and free recall, i.e., remembers the utility from previously searched options or can costlessly revisit them.

At a point in the search process, suppose that the decision maker has opened a set S_i of boxes, which revealed a maximum reward value of $u_i^* = \max\limits_{j \in S_i \cup 0} u_{ij}$, and $\bar{S}_i$ unopened boxes can still be opened. The decision maker now has to decide whether to stop opening boxes, in which case she gets payoff u_i^*, or to continue opening boxes, in which case she needs to decide which unopened box to open next. This decision problem constitutes a dynamic programming problem described by the following Bellman equation:

$$V(\bar{S}_i, u_i^*) = \max\{u_i^*, \max_{j \in \bar{S}_i}\{-c_{ij} + W_j(\bar{S}_i, u_i^*)\}\} \tag{1}$$

where $W_j(\bar{S}_i, u_i^*)$ represents the expected value of continuing to open boxes, i.e., to search, and is defined as

$$W_j(\bar{S}_i, u_i^*) = V(\bar{S}_i \setminus j, u_i^*) \int_{-\infty}^{u_i^*} dF_{ij}(u) + \int_{u_i^*}^{\infty} V(\bar{S}_i \setminus j, u) dF_{ij}(u). \tag{2}$$

In words, given the state space $(\bar{S}_i, u_i^*)$, the decision maker can stop searching and obtain payoff u_i^* or she can continue searching, having to choose which box $j \in \bar{S}_i$ to open next. The box opened next, j, should maximize the expected payoff from continuing to search, net of search costs. In equation (1), this is denoted by $\max_{j \in \bar{S}_i}\{-c_{ij} + W_j(\bar{S}_i, u_i^*)\}$. The value of continuing to search is displayed in equation (2) and has two components: $V(\bar{S}_i \setminus j, u_i^*) \int_{-\infty}^{u_i^*} dF_{ij}(u)$ denotes the decision maker's expected reward from opening box j and finding a reward lower than the best option searched so far, and $\int_{u_i^*}^{\infty} V(\bar{S}_i \setminus j, u) dF_{ij}(u)$ describes her expected payoff from finding a reward higher than the maximum reward observed before searching j.

Economists and marketers have adopted this framework by thinking about consumers as decision makers, products as boxes, and utilities as rewards. Consumers are uncertain about product utilities and have to engage in costly search to resolve their uncertainty.

2.2 The Weitzman (1979) Framework

In most empirical work using sequential search models, a set of optimal decision rules, derived by Weitzman (1979), is used to describe consumer behavior and to derive search and choice probabilities. In this section, we lay out the assumptions required to derive these decision rules, summarize the decision rules, and outline scenarios under which these rules no longer describe optimal search and choice behavior.

2.2.1 Assumptions About the Information Environment

Weitzman (1979) makes the following assumptions regarding the information environment of the sequential search model presented above:

1. Consumers know the true distribution(s) $F_{ij}(u)$ ("rational expectations assumption").

2. Search fully reveals the utility associated with product j.

3. For each consumer i, u_{ij} is independently (across j) drawn from $F_{ij}(u)$.

The first assumption states that consumer beliefs coincide with the true utility distribution. This assumption is typical in many dynamic decision settings of which consumer search is a special case. It is made since consumer beliefs are typically not observed by the researcher, so the assumption allows for an internally consistent model and facilitates empirical work.[2] This assumption rules out learning about the utility distribution (e.g., about prices in the market or other features that affect utility and that consumers are searching for) and is most appropriate in settings in which consumers have experience with a product.[3] The second assumption states that all utility-relevant information about the product is revealed to the consumer upon product inspection. This assumption separates search from the consumer learning literature in which consumers receive an imperfect signal about the true utility each time they consume the product.[4] The third assumption requires that the consumer only obtains information about one product when searching, ruling out the possibility that the outcome of searching one product affects the expected payoff from searching any other product.[5] The third assumption allows us to simplify the dynamic programming problem from equation (1) because it implies that the outcome of searching a particular product does not affect the expected outcome of searching other products. This separability of search decisions is key to deriving the optimal decision rules, which we describe next.

2.2.2 Optimal Decision Rules

Given the assumptions outlined above, we can characterize consumers' optimal search and choice strategies by a set of three decision rules (Weitzman 1979). These decision rules depend on a value called "reservation utility". The reservation utility of a product z_{ij} is the utility level that makes the

[2]For papers that evaluate whether consumers know the true distribution, that determine how quickly consumers learn the true distribution, or that investigate the consequences of assuming that consumer beliefs coincide with the true utility distribution, see Matsumoto and Spence (2016), Ching et al. (2020), and Jindal and Aribarg (2021).

[3]If this assumption is not appropriate, a model of consumer search with learning of the distribution should be estimated instead (see, e.g., Häubl et al. 2010, Koulayev 2013, De los Santos et al. 2017).

[4]For prior work on search that relaxes the assumption that all utility-relevant information is revealed in one search see, e.g., Branco et al. (2012, 2016); Ke et al. (2016); Gardete and Hunter (2020); Ke and Villas-Boas (2019); Ursu et al. (2020); Chick and Frazier (2012); Dukes and Liu (2015); Ursu et al. (2021b).

[5]For models of search with learning across products, see Gardete and Hunter (2020); Hodgson and Lewis (2020).

consumer indifferent between searching product j and receiving z_{ij} with certainty. Mathematically, it equates the marginal gain from searching product j with the marginal cost of doing so, i.e.,

$$\int_{z_{ij}}^{\infty} (u_{ij} - z_{ij}) dF_{ij}(u_{ij}) = c_{ij}. \tag{3}$$

Without loss of generality, we index products in descending order of reservation utilities:

$$z_{i1} \geq z_{i2} \geq \cdots \geq z_{iJ}. \tag{4}$$

With this definition in hand, we now turn to describing the optimal search and choice decision rules.

1. ***Selection Rule****: The consumer searches in decreasing order of reservation utilities.*

   The consumer ranks products in decreasing order of their reservation utilities and proceeds to search them in that order. Therefore, the search order is identical to the ordering of reservation utilities, i.e., the product with the highest reservation utility is searched first, etc.

2. ***Stopping Rule****: Search terminates when the maximum observed utility exceeds the reservation utility of any unsearched product.*

   After searching and learning the realization of post-search utility for a given product, the consumer needs to decide whether to continue searching or to stop. The consumer continues searching if and only if the maximum realized utility among searched products (incl. the outside option $j = 0$) is lower than the maximum reservation utility among unsearched options. If the consumer decides to search product j, it must be that the maximum realized utility among the products searched up to this point $0, 1, \ldots, (j-1)$ is lower than the reservation utility of product j. Formally, the following condition needs to hold for the consumer to search product j:

$$\max\{u_{i0}, u_{i1}, \ldots, u_{i(j-1)}\} \leq z_{ij}. \tag{5}$$

   Otherwise, if there exists no such product, the consumer stops searching. The consumer will also stop if she has searched all available options.

3. ***Choice Rule****: Once the consumer stops searching, she chooses the product with the highest observed utility among all searched options.*

   The consumer chooses product k from the set of searched products $S_i \cup \{0\}$ if product k has the highest utility among the searched products:

$$u_{ik} \geq \max_{j \in S_i \cup \{0\}} u_{ij}. \tag{6}$$

Taken together, equations (4), (5), and (6) describe the order of search, the stopping decision, and the consumer's choice. These equations fully characterize the search and purchase process.

2.3 Parametrizations for Empirical Work

In most empirical work, we assume consumer i's utility from product j has two additively separable components:

$$\begin{aligned} u_{ij} &= \delta_{ij} + \varepsilon_{ij} \\ &= (\xi_{ij} + \mu_{ij}) + \varepsilon_{ij}, \end{aligned} \tag{7}$$

where δ_{ij} denotes the part of utility which is known by the consumer prior to search ("pre-search utility" in the following) and ε_{ij} denotes the part of utility that is only known by the consumer after search ("post-search taste shock" in the following). We assume that the pre-search utility δ_{ij} consists of a component ξ_{ij} that can be observed by the researcher and a taste shock μ_{ij} that cannot be observed by the researcher. In the following, we refer to μ_{ij} as the pre-search taste shock. Both the pre- and post-search taste shocks, μ_{ij} and ε_{ij}, are assumed to be i.i.d. normal with mean zero and standard deviations of σ_μ and σ_ε, respectively.[6] In many settings, one needs to further normalize their variance by setting σ_μ and σ_ε equal to one. We discuss when these additional normalizations are necessary in Section 3, but maintain the more general notation throughout this section.

The additive nature of the utility function allows us to write consumer i's reservation utility as

$$z_{ij} = \delta_{ij} + g(c_{ij}), \tag{8}$$

where $g(c_{ij})$ is a known function that monotonically decreases in search cost c_{ij}, ranges from negative infinity (if search costs go to infinity) to infinity (if search costs are zero), and only depends on search costs and the distribution of the post-search taste shock ε_{ij}.[7] In other words, the reservation utility is equal to the pre-search utility plus an additive term that captures search costs and expected search benefits (via the distribution of post-search taste shocks).

Under the assumption of normally distributed post-search taste shocks, we can derive the following expression for the reservation utility:

$$z_{ij} = \delta_{ij} + m\left(\frac{c_{ij}}{\sigma_\varepsilon}\right) \times \sigma_\varepsilon, \tag{9}$$

[6] It is possible to make different assumptions regarding the distributions of μ_{ij} and ε_{ij}. For example, Elberg et al. (2018) assume a logistic instead of a normal distribution for the post-search error term. For simplicity, we focus on the normal case since it is the most common one, but most of our analysis does not depend on this particular functional form assumption. To ensure that the reservation utility exists and is unique for any value of c_{ij}, we require the mild assumption that the distribution of ε_{ij} is continuous and has full support.

[7] Equation (8) follows from the definition of the reservation utility in equation (3). Changing the variable of integration to $\varepsilon_{ij} = u_{ij} - \delta_{ij}$ yields the following modified reservation utility expression: $\int_{(z_{ij}-\delta_{ij})}^{\infty} (\varepsilon_{ij} - (z_{ij} - \delta_{ij})) dF_{ij}(\varepsilon_{ij}) = c_{ij}$. It follows that $z_{ij} - \delta_{ij} = g(c_{ij})$ where $g()$ denotes the function that assigns the value of $(z_{ij} - \delta_{ij})$ that solves the modified resevervation utility expression for a given value of search costs.

where $g(c_{ij}) = m\left(\frac{c_{ij}}{\sigma_\varepsilon}\right) \times \sigma_\varepsilon$. As discussed above, $g(c_{ij})$ is a function of search costs and expected search benefits. Given the normality assumption for the post-search taste shocks, the expected search benefits are captured by the standard deviation of the post-search taste shocks. The value $m(c_{ij}/\sigma_\varepsilon)$ is the implicit function that solves the following equation (see Kim et al. 2010):

$$\frac{c_{ij}}{\sigma_\varepsilon} = \phi(m) + m \times [\Phi(m) - 1] \tag{10}$$

with ϕ and Φ denoting the standard normal pdf and cdf, respectively. The mapping in equation (10) from search costs and the post-search error variance to reservation utilities will be important when estimating search models which require the researcher to compute reservation utilities for different values of preference and search cost parameters.

Both preferences and search costs are typically parametrized as functions of observable product characteristics:

$$\begin{aligned} \xi_{ij} &= \boldsymbol{X}_j'\boldsymbol{\beta}_i - \alpha_i p_{ij} \\ c_{ij} &= \boldsymbol{Z}_j'\boldsymbol{\gamma}_i \end{aligned}$$

where p_{ij} denotes price and $\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$ denote vectors of product characteristics that enter preference and/or search costs and may vary across consumers. Both sets of variables can include product fixed effects, physical product characteristics, and variables that capture saliency such as product ranking on a webpage or advertising. Search costs can include an intercept with a consumer-specific coefficient.[8] There could be overlap in the variables entering utility and search costs and we discuss the separate identification of search costs and preferences in Section 4. We assume that non-price characteristics ($\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$) do not vary across customers, while prices p_{ij} are consumer-specific. The typical setting we have in mind is one where different consumers are observed at different points in time and therefore face different prices, whereas $\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$ denote characteristics such as physical attributes that do not vary over time. This simplification is entirely expositional and it is easy to allow for other variables to also vary across consumers.

This utility specification is analogous to the way utility is typically specified in full information settings. It nests a model without preference heterogeneity (in which case we can drop the i subscript from preference parameters) as well as a model with observed and/or unobserved heterogeneity in preferences. In many applications, search costs are assumed to vary across individuals, but not across products. More specifically, search costs are often assumed to be drawn from a log-normal distribution (e.g., Jiang et al., 2021, Morozov et al., 2021). Such a specification is most appropriate when we think of search costs as opportunity costs of time which are unlikely to differ across products in many settings. Recent work has modeled search costs as a function of variables that influence salience such as product rankings (Ursu 2018) or advertising (Ursu et al. 2021a),

[8] In empirical work, search costs are frequently constrained to be positive by specifying the function of parameters as an exponential.

in which case search costs tend to vary across products. We explicitly allow for the possibility of search cost differences across consumers and products when discussing identification.

Finally, we note that consumer-level search data can differ with regards to whether choices of the outside option are observed, i.e., whether the data contain information on consumers who did not search (and consequently did not purchase) and consumers who searched but did not make a purchase. If the data are conditional on searching it is common to assume that the first search is free in order to rationalize that each search spell contains at least one product.[9] If consumers that could have searched but decided not to are also observed, it is not necessary to make the assumption of a costless first search. If the data only contain searches that end in a purchase, the choice of the outside option does not need to be modeled (see, e.g., Morozov et al. 2021). If the data also contain observations in which consumers searched but did not purchase, an outside option (with deterministic utility set equal to zero) is typically included in the model and the outside option is assumed to always be available, i.e., no search costs need to be incurred to reveal the utility of the outside option.

2.3.1 Relation to Price Search Models

For most of the discussion in this and the following sections, we focus on the case in which consumers search to learn about a taste shock ε_{ij}, sometimes also referred to as a "match value," that is unobserved by the researcher pre- and post-search. Alternatively, in specific empirical settings, one could model consumers as searching over a specific product characteristic such as price (Honka, 2014, Honka and Chintagunta, 2017). In such a case, it is assumed that consumers know the distribution of this characteristic, but not the realization for a given product. Most of our analysis is unaffected by changing the object consumers are searching over with three exceptions: first, the post-search error term ε_{ij} is replaced by the characteristic the consumer is searching over. Therefore, the utility function only contains the pre-search error term μ_{ij} but not ε_{ij}. Second, when searching over an observable product characteristic, the distribution of post-search utility is typically known by the researcher, i.e., it is given by the observed empirical distribution of post-search utility. Prior work has typically assumed a functional form for the post-search utility and estimated the parameters of this distribution (such as the mean and variance of a normal distribution) from data. In contrast, in a match-value search model, the researcher makes assumptions on both the functional form for the post-utility distribution and its parameters. And lastly, the realized value of the specific product characteristic is usually observed by the researcher post-search.[10]

More generally, price and match value search models are a subset of a larger class of models with an additively separable utility function $u_{ij} = (\boldsymbol{X}_j'\boldsymbol{\beta}_i + \mu_{ij}) + (\boldsymbol{L}_j'\boldsymbol{\kappa}_i + \varepsilon_{ij})$, in which the terms in the first brackets are known prior to search and consumers learn about the terms in the second brackets through search. Match value models reduce the post-search part to ε_{ij}, whereas price

[9] Alternatively, Reinganum (1979) assumes that search costs are so low that it is optimal for all consumers to search at least once.

[10] These three differences have consequences for the taste shocks normalizations (see Section 3.3) as well as the elasticy and welfare expressions which have to be adjusted for the case of price search.

search models reduce the post-search part to price multiplied by consumer preference for price. While less common in empirical work, the post-search part of utility could, in principle, be based on multiple characteristics as well as a combination of observed and unobserved factors (see Yao et al. 2017, Abaluck et al., 2022, Compiani et al., 2022, Moraga-González et al., 2022).

3 Taste Shocks Normalizations

The utility function specified in equation (7) contains two idiosyncratic taste shocks: μ_{ij} and ε_{ij}. The first taste shock μ_{ij} is part of the pre-search utility, while the second taste shock ε_{ij} is revealed after search. Here, we discuss why the model contains two error terms, when we need to normalize their variances, and the consequences of these normalizations for the interpretation of other model parameters.

3.1 The Role of the Two Taste Shocks in the Model

The post-search error term ε_{ij} is the object that consumers need to search over to learn its realization. Without such an error term, there would be no need to model search and instead the utility function described in equation (7) (minus the post-search part ε_{ij}) would give rise to a perfect information demand model.

The less obvious question is why μ_{ij}, a second error term that consumers observe before engaging in search, is needed. Given the reservation utility expression in equation (8), search costs c_{ij}, and the utility specification from equation (7), note that reservation utilities can be written as

$$z_{ij} = \boldsymbol{X}_j' \boldsymbol{\beta}_i - \alpha_i p_{ij} + \mu_{ij} + g(c_{ij}), \tag{11}$$

where $\boldsymbol{X}_j$, p_{ij}, and $g(c_{ij})$ are observed/ estimated by the researcher. Without μ_{ij} in the model, the search order would be a deterministic function of observed variables and model parameters. Therefore, the pre-search error μ_{ij} introduces a stochastic element which allows any search order to occur with positive probability.[11]

3.2 Normalization of the Taste Shock Variances

Next, we discuss why the variances of both errors terms are often normalized when taking the model to data. Recall that μ_{ij} and ε_{ij} are i.i.d normally distributed with mean zero and standard deviations of σ_μ and σ_ε, respectively; thus, $\mu_{ij} = \sigma_\mu \bar{\mu}_{ij}$ and $\varepsilon_{ij} = \sigma_\varepsilon \bar{\varepsilon}_{ij}$. The bar notation denotes the standard normal draws corresponding to each type of taste shock. With this notation in hand, we can re-write consumer utility to make the role of the standard deviations of the two error terms explicit:

[11] In some settings, the nature of the data might allow the researcher to omit the pre-search error term. For example, in Ursu (2018) the search order is unobserved, so the estimated model omits the selection rule. Choi and Mela (2019) allow for an error term that enters search costs, thus introducing a stochastic element into equation (11) via $g(c_{ij})$. Chung et al. (2019) similarly allow for search costs to vary at the consumer/product level.

$$u_{ij} = \boldsymbol{X}'_j\boldsymbol{\beta}_i - \alpha_i p_{ij} + \sigma_\mu \bar{\mu}_{ij} + \sigma_\varepsilon \bar{\varepsilon}_{ij}.$$

We then divide both utility and search costs by the standard deviation σ_μ of the pre-search shocks to obtain:

$$\begin{aligned}
\tilde{u}_{ij} &= \boldsymbol{X}'_j(\boldsymbol{\beta}_i/\sigma_\mu) - (\alpha_i/\sigma_\mu)p_{ij} + \bar{\mu}_{ij} + (\sigma_\varepsilon/\sigma_\mu)\bar{\varepsilon}_{ij} \\
&= \boldsymbol{X}'_j\tilde{\boldsymbol{\beta}}_i - \tilde{\alpha}_i p_{ij} + \bar{\mu}_{ij} + \tilde{\sigma}_\varepsilon \bar{\varepsilon}_{ij}, \\
\tilde{c}_{ij} &= c_{ij}/\sigma_\mu.
\end{aligned}$$

Here, we use the tilde notation for all variables scaled by σ_μ (e.g., $\tilde{u}_{ij} = u_{ij}/\sigma_\mu$). Reservation utilities are also re-scaled in the same way as utilities. This can be shown by considering the reservation utility expression in equation (9). When dividing utility and search costs by σ_μ, reservation utilities become

$$\begin{aligned}
\tilde{z}_{ij} &= \boldsymbol{X}'_j(\boldsymbol{\beta}_i/\sigma_\mu) - (\alpha_i/\sigma_\mu)p_{ij} + \bar{\mu}_{ij} + m(c_{ij}/\sigma_\varepsilon) \times (\sigma_\varepsilon/\sigma_\mu) \\
&= z_{ij}/\sigma_\mu.
\end{aligned}$$

Note that both search costs and the standard deviation of the post-search error are divided by σ_μ and therefore σ_μ cancels out of the expression inside the function $m(\cdot)$.

Since this transformation rescales utilities, search costs, and reservation utilities in the same way, it does not alter the search order, the stopping, or the purchase decisions. Thus, we can set $\sigma_\mu = 1$ without loss of generality. This normalization is analogous to scale normalizations in demand models without search frictions, such as the standard logit and probit models of demand (see Maddala 1983, Cameron and Heckman 1998, Breen et al. 2018).

Now consider estimating the standard deviation of post-search shocks σ_ε. Although search costs and post-search shocks are not fully co-linear, it is hard to estimate both terms in practice. This is the case because, intuitively, increasing the post-search shock standard deviation σ_ε and decreasing search costs both lead to more search. Thus, in practice, only the magnitude of search costs *relative* to the post-search shock standard deviation σ_ε can be estimated. Yavorsky et al. (2021) provide an extensive set of simulations showing that it is difficult to separately estimate search costs and the post-search shock standard deviation. Similarly, Morozov et al. (2021) report results suggesting that the standard deviation of the post-search taste shock is difficult to estimate.[12] One exception is Morozov (2022) who estimates the variance of one of the error terms.[13] He argues that a larger role of characteristics on search versus purchase decisions identifies the variance of one of the taste shocks because purchase decisions are affected by both taste shock realizations, while search decisions are only affected by the pre-search taste shock. Therefore, although there are some

[12]See endote 16 in Morozov et al. 2021.

[13]Morozov (2022) sets the variance of the post-search taste shock equal to one and estimates the standard deviation of the pre-search error term.

mixed results regarding the feasibility of estimating one of the taste shock standard deviations, it appears that there is often limited variation in the data to allow researchers to separately estimate search costs and the post-search error term standard deviation. Hence, σ_ε is set to one in many empirical applications (e.g., Chen and Yao, 2017; Morozov et al., 2021; Ursu et al., 2021a).

There are two approaches to avoiding the post-search taste shock normalization in the recent empirical literature. First, in cases in which consumers search over price or other observed characteristics (e.g., Honka, 2014; Honka and Chintagunta, 2017), the researcher can use the empirical distribution of this characteristic to compute the variance of the post-search utility; thus no additional assumptions are required. Second, in cases in which the researcher has access to additional data, e.g., a variable that affects search costs, but not post-search utility, she can use such search cost shifters to separately estimate the magnitude of search costs and the variance of the post-search utility. For example, Yavorsky et al. (2021) estimate a search model in the context of car dealership visits and argue that distance to the dealership affects search costs, but not post-search utility. This search cost shifter allows them to estimate the post-search utility variance. In other settings, variables such as product rankings could also be argued to shift search costs but not post-search utility (Ursu 2018). Given the importance of monetizing search costs and running counterfactuals that alter search costs (see the discussion in the next sub-section), such search cost shifters are likely to be crucial for many empirical studies of search behavior.

3.3 Monetizing Search Costs and Counterfactuals

An important consequence of setting both standard deviations to one is that search cost estimates cannot be expressed in monetary terms by dividing them by the price coefficient. As discussed in the previous subsection, search costs can only be estimated relative to the post-search error standard deviation. Therefore, the estimated search cost in the re-scaled model is given by

$$\frac{\tilde{c}_{ij}}{\tilde{\sigma}_\varepsilon} = \frac{c_{ij}/\sigma_\mu}{\sigma_\varepsilon/\sigma_\mu} = \frac{c_{ij}}{\sigma_\varepsilon}.$$

Thus, the ratio of the estimated search cost to the estimated price coefficient equals the ratio of α_i/σ_μ to $c_{ij}/\sigma_\varepsilon$ and is not equivalent to the monetary value of search costs. However, σ_ε only affects estimated search costs but not utility parameters known to the consumer prior to searching, such as the price coefficient and product intercepts. Therefore, while we cannot monetize search costs, we can still monetize any parameter that is known to consumers prior to search (see Morozov et al. 2021).

We believe that the consequences of the two normalizations discussed above are not well understood in empirical work and might lead to misleading search cost estimates. The role of the normalizations is particularly important because in many empirical papers consumers are observed to search relatively few products which translates into large estimated search costs. However, a lack of search can be driven by either high search costs or a low benefit of search. When fixing the benefit of search by fixing the standard deviation of the post-search taste shock, only search costs are allowed to rationalize short search spells. Given the discussion outlined above, one should therefore be careful in interpreting monetized search cost estimates when the standard deviation of the post-search taste shock is set to a fixed value.

Similarly, counterfactuals that alter search costs are directly affected by the post-search taste shock normalization. For example, a researcher might be interested in how consumers' choices change when removing search frictions. Reducing search costs to zero gives all consumers free access to the post-search taste shocks. Because consumers' choices are affected by the variance of the post-search utility component when search costs are removed, counterfactual choice behavior as well as the welfare consequences of removing search costs will depend on the arbitrarily set value of the post-search taste shock standard deviation. Therefore, counterfactuals that alter search costs should not be considered when the estimated model involves the normalization of both standard deviations.

4 Identification

In this section, we provide a discussion of identification that we split into two parts. We first provide formal identification proofs, which show that different values of the parameters imply different distributions of observable data (Andrews et al. 2017; Matzkin 2013; Hsiao 1983). Second, we provide a more "informal" discussion of identification which is commonly used in applied economics and marketing papers. This informal discussion verbally describes which variation in the observables/ which moments in the data (together with distributional and functional form assumptions, etc.) are informative about specific model parameters. Our formal identification arguments are based on subsets of the observed data (such as the probability of a product being search first), whereas the informal arguments describe more exhaustively what variation determines a particular parameter estimate. For lack of a better term, we continue to use the term "informal identification" below, but previous literature has also referred to it as "parameter sensitivity" (Andrews et al. 2017). Most papers in the consumer search literature use the term identification to refer to this more informal discussion (e.g., Kim et al. 2010; Seiler 2013; Honka 2014; Ursu 2018). However, we believe that both types of arguments fulfill complementary roles and enhance the researcher's understanding of the sequential search model.

A challenge in presenting a general identification discussion is that identification depends on the characteristics of the data and often also on institutional details of the empirical application. In this paper, we focus on identification and estimation of the sequential search model with individual-level data and abstract from idiosyncrasies of specific empirical applications. The data we have in mind when discussing identification contain information on consumers' search order, on the set of products searched and those not searched, and on purchase decisions. A typical example of such a data set is online browsing data. We focus on this type of data because we view it as the most common data that is and will be available to researchers. We also assume throughout this section that each consumer's first search is free and that there is no outside option.[14]

The main identification challenge in the sequential search model consists of separately identifying consumer preference parameters from search costs. Using the parametrization of the model we introduced in Section 2.3 as well as the assumption of standard normally distributed pre- and post-search taste shocks,[15] our goal is to show how parameters (α_i, β_i), which affect consumer utility, are separately identified from search costs, parameterized by γ_i. Individual-level data on search order, number of searches, and purchases made, together with the search rules presented in Section 2.2.2 constitute the necessary inputs for our identification arguments.

4.1 Specification A: Homogeneous Model with Common Search Costs

We start with a formal identification discussion of the frequently used version of the sequential search model in which preferences are homogeneous across consumers and search costs are constant across products and consumers. Hence, utility is given by

$$u_{ij} = X_j'\boldsymbol{\beta} - \alpha p_{ij} + \mu_{ij} + \varepsilon_{ij}$$

and preference parameters $(\alpha, \boldsymbol{\beta})$ do not have consumer-specific subscripts. Search costs are common across consumers and products, i.e., $c_{ij} = c$. Given this model specification, in the following, we show that we are able to separate decisions that are driven by preference parameters from decisions that are driven by both preferences and search costs. We first turn to the identification of preference parameters.

4.1.1 Search Order and Preference Parameters

In a model with common search costs, reservation utilities for all products depend on search costs via the common function $g(c)$. Because $g(c)$ does not differ across products, it does not affect the ordering of reservation utilities and therefore, search order is solely determined by preferences. We can derive the following probability that product k is searched first:

$$\begin{aligned}
Pr(\text{product } k \text{ searched first}) &= Pr(z_{ik} \geq z_{ij} \quad \forall j) \\
&= Pr(X_k'\boldsymbol{\beta} - \alpha p_{ik} + \mu_{ik} + g(c) \geq X_j'\boldsymbol{\beta} - \alpha p_{ij} + \mu_{ij} + g(c) \quad \forall j) \\
&= Pr(X_k'\boldsymbol{\beta} - \alpha p_{ik} + \mu_{ik} \geq X_j'\boldsymbol{\beta} - \alpha p_{ij} + \mu_{ij} \quad \forall j) \\
&= \int \mathbf{1}\left(X_k'\boldsymbol{\beta} - \alpha p_{ik} + \mu_{ik} \geq X_j'\boldsymbol{\beta} - \alpha p_{ij} + \mu_{ij} \quad \forall j\right) \phi\left(\boldsymbol{\mu}\right) d\boldsymbol{\mu}
\end{aligned}$$

where $\boldsymbol{\mu}$ denotes the vector of pre-search taste shocks for all products.

[14] This assumption is made for the purpose of an easier exposition. It is easy to modify the identification arguments to include an outside option.

[15] We note that the identification arguments below only require continuous and full support for the error terms, but do not directly rely on the assumption of normality.

The third line follows from the fact that reservation utilities for all products involve the same additive term $g(c)$ and hence this term is irrelevant for the ranking of reservation utilities across products. The forth line follows under the assumption that μ_{ij} is standard normally distributed; under this assumption, the first search probabilities are given by standard probit expressions. First searches therefore provide us with a similar expression as purchases in a full information setting and identification arguments with regard to preference parameters carry over directly. For example, the price coefficient is identified by the extent to which products with lower prices are more likely to be searched first.

4.1.2 Stopping Decisions and Search Costs

Conditional on preference parameters, we can identify search costs based on consumers' stopping decisions. In particular, stopping probabilities after the first search are given by:

$$\begin{aligned}
& Pr(\text{stop after 1st search}|\text{product } k \text{ searched first}) \\
& Pr(u_{ik} \geq z_{ij} \quad \forall j \neq k \mid z_{ik} \geq z_{ij} \quad \forall j \neq k) \\
= \quad & \frac{Pr(u_{ik} \geq z_{ij} \text{ and } z_{ik} \geq z_{ij} \quad \forall j \neq k)}{Pr(z_{ik} \geq z_{ij} \quad \forall j \neq k)},
\end{aligned}$$

where the expression in the denominator only depends on preference parameters (because reservation utilities only depend on preference parameters), which are identified from the search order as discussed above.

We can re-write the numerator as follows:

$$\begin{aligned}
& Pr(u_{ik} \geq z_{ij} \text{ and } z_{ik} \geq z_{ij} \quad \forall j \neq k) \\
= \quad & Pr(X_k'\boldsymbol{\beta} - \alpha p_{ik} + \varepsilon_{ik} + \mu_{ik} \geq X_j'\boldsymbol{\beta} - \alpha p_{ij} + \mu_{ij} + g(c) \text{ and } z_{ik} \geq z_{ij} \quad \forall j \neq k) \\
= \quad & Pr(X_k'\boldsymbol{\beta} - \alpha p_{ik} + \varepsilon_{ik} - g(c) + \mu_{ik} \geq X_j'\boldsymbol{\beta} - \alpha p_{ij} + \mu_{ij} \text{ and } z_{ik} \geq z_{ij} \quad \forall j \neq k).
\end{aligned}$$

We established above that preference parameters $(\alpha, \boldsymbol{\beta})$ are identified based on the first search probabilities. Because the distributions of the taste shocks ε_{ij} and μ_{ik} are known, the only free parameter is the search cost parameter which enters the equation above via $g(c)$. Because $g(c)$ is monotonically decreasing in search costs and ranges from minus infinity to plus infinity, there exists a unique value of search costs (conditional on preference parameters) that rationalizes the stopping probability observed in the data.[16] Intuitively, higher search costs lead to a higher probability of stopping after the first search, because they lower the value of reservation utilities and hence the

[16] In more detail, for $g(c) \rightarrow +\infty$ the numerator (and therefore the entire expression) goes towards zero. Instead, for $g(c) \rightarrow -\infty$ the numerator goes towards $Pr(z_{ik} \geq z_{ij} \quad \forall j \neq k)$ because the first condition is always fulfilled and hence the whole expression goes towards one. Increasing search costs monotonically increases the stopping probability.

condition that the maximum realized utility of the first product exceeds the maximum reservation utility among remaining products is more likely to be fulfilled.

4.1.3 Informal Discussion

Our identification argument above is based on the identities of products searched first and the probability of stopping after the first search. We now broaden the focus to a more informal discussion of which moments of the data provide information on preference and search cost parameters.

With regards to the search order, data on the second, third, etc. search in the search sequence contains additional information that also helps estimate preference parameters. The probability that a specific product is searched second, third, etc. is given by a similar probit expression as first search probabilities, except for the fact that reservation utilities at every stage of the search process only need to be larger than the reservation utilities of products not yet searched. The probability that a consumer searched products S_i in order $\{1, ..., H_i\}$ (where we index products by the order in which they are searched) is given by

$$\begin{aligned} Pr(\text{order } \{1,...,H_i\}) &= Pr(z_{i1} \geq z_{ij} \quad \forall j) \times Pr(z_{i2} \geq z_{ij} \quad \forall j \neq 1) \\ & \quad ... \times Pr(z_{iH_i} \geq z_{ij} \quad \forall j \notin S_i). \end{aligned}$$

This formula contains the first search probability as its first term, but also depends on the search probabilities beyond the first search. The expression above suggests that the extent to which products with specific characteristics are searched earlier in the search process provides information to help estimate preference parameters. As was the case for the first search probabilities, the expression that describes the probability of a particular search sequence occurring does not involve search costs (because consumers search products in decreasing order of pre-search utility) and hence search order only provides information about preference parameters.

We can similarly write down expressions for the probability of stopping after the second, third, etc. search. These expressions contain information on consumer search costs, because a higher search cost will lead to a higher probability of ending the search process after every search. Therefore, continuation and stopping decisions after each search provide additional variation that informs the estimated value of search costs.

Finally, purchase decisions in a search model also provide information that helps estimate preference parameters. It is easiest to see this by analyzing the purchase probability expression we present later in Section 5.1. It shows that purchase probabilities take a probit form with an additional truncated error term, where the truncation depends on search costs. Conditional on search costs, preference parameters influence the degree to which products with certain characteristics are more likely to be purchased.

4.2 Specification B: Homogeneous Model with Product-Specific Search Cost

Next, we consider a model in which search costs are product-specific. Such a model specification might be appropriate if factors such as webpage rankings or the salience of a product on the store shelf or on a webpage are likely to affect the cost of searching products independent of consumers' preferences. We consider the case where search costs differ across products in a fully flexible manner, but one could also parameterize search costs as a function of observed characteristics such as product rankings.

We again proceed in two steps similar to the arguments presented for Specification A. However, in this setting, the search order does not solely depend on preferences. Instead, we show that a component of the reservation utility is also identified from the observed search order. In a second step, we discuss how the preference and search cost components of the reservation utilities can be identified from consumers' continuation and stopping decisions.

4.2.1 Inferring Reservation Utilities

According to the selection rule, consumer i searches products in order of decreasing reservation utilities z_{ij}. Recall from equation (11) that the reservation utility can be written as

$$\begin{aligned} z_{ij} &= X_j'\boldsymbol{\beta} - \alpha p_{ij} + g(c_j) + \mu_{ij} \\ &= \bar{z}_j - \alpha p_{ij} + \mu_{ij} \end{aligned}$$

where $\bar{z}_j = X_j'\boldsymbol{\beta} + g(c_j)$. The decision to search product k first can be expressed as

$$\begin{aligned} Pr(\text{product } k \text{ searched first}) &= Pr(z_{ik} \geq z_{ij} \quad \forall j) \\ &= Pr(\bar{z}_k - \alpha p_{ik} + \mu_{ik} \geq \bar{z}_j - \alpha p_{ij} + \mu_{ij} \quad \forall j) \end{aligned}$$

Similar to the arguments made for Specification A, this expression is a standard probit expression and allows us to identify α and $\bar{\boldsymbol{z}} = (\bar{z}_1, ..., \bar{z}_J)$. However, we cannot separately identify $\boldsymbol{\beta}$ from c_j; we can only identify $\bar{z}_j$ which is a function of both $\boldsymbol{\beta}$ and c_j.

4.2.2 Separating Preferences from Search Costs

Next, we turn to stopping patterns to decompose $\bar{\boldsymbol{z}}$ into its preference and search cost components. The probability that, conditional on searching product k first, the consumer stops after the first search is given by the same expression as in Specification A:

$$
\begin{aligned}
& Pr(\text{stop after 1st search}|\text{product } k \text{ searched first}) \\
= \quad & \frac{Pr(u_{ik} \geq z_{ij} \text{ and } z_{ik} \geq z_{ij} \quad \forall j \neq k)}{Pr(z_{ik} \geq z_{ij} \quad \forall j \neq k)}.
\end{aligned}
$$

The expression in the denominator depends on $\bar{\boldsymbol{z}}$ and α, which are identified from the search order.

We can re-write the numerator as follows:

$$
\begin{aligned}
& Pr(u_{ik} \geq z_{ij} \text{ and } z_{ik} \geq z_{ij} \quad \forall j \neq k) \\
= \quad & Pr(X_k'\beta - \alpha p_{ik} + \varepsilon_{ik} + \mu_{ik} \geq \bar{z}_j - \alpha p_{ij} + \mu_{ij} \text{ and } z_{ik} \geq z_{ij} \quad \forall j \neq k) \\
= \quad & Pr(\bar{z}_k - \alpha p_{ik} - g(c_j) + \varepsilon_{ik} + \mu_{ik} \geq \bar{z}_j - \alpha p_{ij} + \mu_{ij} \text{ and } z_{ik} \geq z_{ij} \quad \forall j \neq k)
\end{aligned}
$$

Because the distributions of the taste shocks ε_{ij} and μ_{ik} are assumed to be known, the only free parameter that enters the equation is the product-specific search cost term $g(c_j)$. Similar to the discussion regarding the homogenous search costs case, there is a unique value of product-specific search costs that rationalizes the observed conditional stopping decision after the first search. We can derive an analogous expression for each product, and hence product-specific stopping probabilities identify $g(c_k)$ and therefore search costs c_k for each product.Note that $\bar{z}_j$ is a linear combination of product-specific search costs and preferences over characteristics $X_j'\boldsymbol{\beta}$. Therefore, it follows that the coefficients $\boldsymbol{\beta}$ are identified because $\bar{z}_j$ and c_j are identified based on the arguments above.

To gain some intuition for how product-specific search costs and preference parameters can be identified, consider product A with high pre-search utility and high product-specific search cost and product B with low pre-search utility and low search costs such that both products have the same reservation value and therefore the same probability of being searched first. Product A will have a higher probability of stopping after it has been searched relative to product B for the following reason: after the search, the search costs for product A are sunk and no longer decision-relevant, but the higher realized utility makes it less attractive for the consumer to continue searching.

We note that one could parameterize search costs as a function of certain product characteristics and identify the degree to which these characteristics impact search costs. Because search costs are separately identified from preferences, these characteristics could overlap with those that enter utility and the researcher can estimate a flexible search model to learn whether a given product characteristic shifts utility and/or search costs. In many cases, it might be natural to include specific variables that increase the salience of a product such as product rankings or advertising solely as part of search costs rather than utility.

4.2.3 Informal Discussion

Similar arguments to those presented for the homogeneous search costs case also apply to the case of product-specific search costs. The full search order beyond only the first search provides additional information to estimate $\bar{\boldsymbol{z}} = (\bar{z}_1, ..., \bar{z}_J)$ and the price coefficient α. Furthermore, continuation and stopping decisions after every search provide variation that helps identify product-specific search costs. Finally, purchase probabilities are determined by both search costs and preferences. Preference parameters influence the sensitivity of purchases to product characteristics and higher product-specific search costs lower the purchase probability for a specific product.

4.3 Specification C: Preference Heterogeneity

Finally, we consider a version of the sequential search model with constant search costs and heterogeneity in preferences, in particular heterogeneous tastes over product characteristics. Thus utility is given by

$$u_{ij} = X_j'\boldsymbol{\beta_i} - \alpha_i p_{ij} + \mu_{ij} + \varepsilon_{ij},$$

where preferences over characteristics $\{X_j, p_{ij}\} = \{x_{1j}, ..., x_{Kj}, p_{ij}\}$ are distributed according to

$$\begin{bmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{iK} \\ \alpha_i \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \\ \vdots \\ \bar{\beta}_K \\ \bar{\alpha} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \sigma_K & 0 \\ 0 & 0 & 0 & 0 & \sigma_p \end{bmatrix} \right).$$

As we show later in Section 5.2, elasticities in a sequential search model take a similar form to those derived from a perfect information model with the same utility function. Therefore, random coefficients on specific product characteristics serve the role of informing the pattern of cross-price elasticities such that products with more similar characteristics tend to have higher cross-price elasticities. In this section, we show that search data viewed through the lens of the sequential search model provides information that is particularly useful for estimating the parameters of a utility specification with random coefficients.

We first re-iterate that first search probabilities take the following form

$$\begin{aligned} Pr(\text{product } k \text{ searched first}) &= Pr(z_{ik} \geq z_{ij} \quad \forall j) \\ &= Pr(X_k'\boldsymbol{\beta_i} - \alpha_i p_{ik} + \mu_{ik} + g(c) \geq X_j'\boldsymbol{\beta_i} - \alpha_i p_{ij} + \mu_{ij} + g(c) \quad \forall j) \\ &= Pr(X_k'\boldsymbol{\beta_i} - \alpha_i p_{ik} + \mu_{ik} \geq X_j'\boldsymbol{\beta_i} - \alpha_i p_{ij} + \mu_{ij} \quad \forall j). \end{aligned}$$

Contrary to the specification with homogenous preferences discussed in Section 4.1, we now need to integrate over the distribution of the error terms and the random coefficients in order to compute the expression in the last line above. Importantly, the expression above does not depend on search costs and takes the same form as the purchase probability expressions from a standard perfect information choice model. Therefore, identification arguments from perfect information choice models such as those presented in Berry and Haile (2014) carry over directly to our setting and thus preference parameters are identified from first search probabilities even in the presence of random coefficients. The identification argument for the identification of search costs conditional on preference parameter is the same as the one presented in Section 4.1.

4.3.1 Informal discussion

Data on the order of search is particularly important to estimate heterogeneity in consumers' tastes. As established above, when search costs do not vary across products, consumers will search products in the order of pre-search utility. Therefore the probability of observing a particular search order is given by

$$\begin{aligned} Pr(\text{order } \{1,...,H_i\}) &= Pr(z_{i1} \geq z_{i2} \geq ... \geq z_{iH_i} \geq z_{il} \quad \forall l \notin S_i) \\ &= Pr(\delta_{i1} \geq \delta_{i2} \geq ... \geq \delta_{iH_i} \geq \delta_{il} \quad \forall l \notin S_i), \end{aligned}$$

where $\delta_{ij} = X_j' \boldsymbol{\beta_i} - \alpha_i p_{ij} + \mu_{ij}$ denotes pre-search utility.

This expression reveals that search order provides information on first-best, second-best, etc. products in terms of their pre-search utility. Search order information therefore plays a role similar to second-choice data obtained from surveys (Berry et al. 2004). Depending on the length of search spells, the search order information might be substantially richer than second-choice data because it can contain information on more than just the first two highest-ranked products.

Intuitively, the similarity of the products being searched in terms of their characteristics is informative about heterogeneity in preferences over characteristics. For example, Morozov (2022) shows in an empirical application to the hard-drive market that consumers who search a solid state hard drive (SSD) are more likely to search other SSDs and consumers that do not search an SSD are more likely to continue searching regular hard drives.[17] More generally, the similarity of products in consumers' search sets in terms of their characteristics is informative about the variance terms that determine the degree of preference heterogeneity over specific characteristics.

5 Model Properties: Elasticities, Welfare, and Counterfactuals

In this section, we derive expressions for own- and cross-price elasticities as well as consumer welfare that take search frictions into account. To derive these expressions, we first describe a reformulation of purchase probabilities implied by the sequential search model that we use to derive

[17] See Section 3.4 and Appendix A6 in Morozov (2022).

elasticities and consumer welfare. We also describe a set of counterfactuals that directly rely on the search model framework, such as a change or removal of search costs. To simplify notation, we focus on a model without heterogeneity in preferences and search costs in this section. When relevant, we discuss the role of unobserved heterogeneity in preferences or search costs.

5.1 Alternative Expression for Purchase Probabilities

A convenient expression for purchase probabilities is provided by Armstrong (2017) and Choi et al. (2018), who show that the sequential search model can be represented as a discrete choice model in which the consumer chooses the product with the highest *effective value* w_{ij}, defined as $w_{ij} = min(u_{ij}, z_{ij})$. Recall that the reservation utility of product j can be written as $z_{ij} = X_j'\boldsymbol{\beta} - \alpha p_{ij} + g(c) + \mu_{ij}$. The effective utility of product k can therefore be expressed as $w_{ij} = X_j'\boldsymbol{\beta} - \alpha p_{ij} + \mu_{ij} + min(\varepsilon_{ij}, g(c))$. We denote the effective utility of the outside option by $w_{i0} = \mu_{i0}$. Using these expressions, we can compute the probability that product k is purchased as

$$
\begin{aligned}
Pr_{ik} &= Pr(w_{ik} \geq w_{ij} \quad \forall j \in J \cup \{0\}) \\
&= Pr(X_k'\boldsymbol{\beta} - \alpha p_{ik} + min(\varepsilon_{ik}, g(c)) + \mu_{ik} \\
&\qquad \geq X_j'\boldsymbol{\beta} - \alpha p_{ij} + min(\varepsilon_{ij}, g(c)) + \mu_{ij} \quad \forall j \in J \cup \{0\}).
\end{aligned}
$$

This expression resembles purchase probabilities in perfect information discrete choice models and only differs due to the inclusion of the truncated distributions of post-search taste shocks ε_{ij}.

To build intuition for this alternative expression for purchase probabilities, it is instructive to consider how consumer search and purchase behavior changes as we increase search costs starting from zero where the consumer searches all options and choices are determined by the full utility including the post-search taste shocks. As we increase search costs, consumers will search fewer products and not discover their post-search taste shocks. Therefore, post-search taste shocks become less relevant for the consumer's purchase decision which is captured by the distribution becoming more truncated (from above) as search costs increase.

Aggregate market-shares conditional on prices $\boldsymbol{p}_i$ are obtained by integrating out the pre- and post-search error terms:

$$
\begin{aligned}
Pr_k(\boldsymbol{p}_i) &= \int_\varepsilon \int_\mu \mathbf{1}(X_k'\boldsymbol{\beta} - \alpha p_{ik} + min(\varepsilon_{ik}, g(c)) + \mu_{ik} \\
&\qquad \geq X_j'\boldsymbol{\beta} - \alpha p_{ij} + min(\varepsilon_{ij}, g(c)) + \mu_{ij} \quad \forall j \in J \cup \{0\}) \\
&\qquad \times \phi(\boldsymbol{\varepsilon})\phi(\boldsymbol{\mu}) d\boldsymbol{\varepsilon} d\boldsymbol{\mu}. \qquad (12)
\end{aligned}
$$

We re-iterate that, for expositional simplicity, we present the case when the only source of stochasticity comes from the two taste shocks. If the model also includes random coefficients that enter preference and/or search costs, the researcher would need to numerically integrate over the distribution of random coefficients as well.

5.2 Elasticities

Based on the derivations above, the own- and cross-price elasticities of demand are given by standard probit elasticities except for the presence of the post-search taste shocks. Importantly, a required assumption for the Weitzman framework is the independence of post-search taste shocks across products (see Sub-section 2.2). Therefore, the sequential search model behaves similarly to a perfect information model with the only difference being an additional (truncated) error term that enters the effective value of each product and is independently distributed across products.

At the extreme, for very large search costs, the impact of the post-search taste shocks ε_{ij} goes to zero and own-price elasticities become standard probit elasticities. When search costs decrease, the post-search errors will play an increasingly important role in choice probabilities because the upper bound that truncates the post-search errors increases when search costs are lower. Intuitively, choices in a setting with finite search costs are partly determined by the post-search taste shocks which are uncorrelated across products. The presence of this additional determinant of choice (on top of the pre-search utility) lessens the impact of price (as well as other characteristics) on choice and therefore lowers the elasticity. This pattern is easiest to see for the extreme case of zero search costs. In this case, choice probabilities are given by probit expression because ε_{ij} enters without an upper bound. Hence, the two taste shocks together $\varepsilon_{ij} + \mu_{ij}$ are normally distributed with a variance that is twice as large as the variance of μ_{ij} alone (due to both variables being standard normally distributed).

A similar logic applies to cross-price elasticities which are also lower in a model with finite search costs where post-search taste shocks impact purchase decisions. Importantly, the post-search taste shocks behave differently from other error components such as random coefficients. Random coefficients lead to a correlation in purchase probabilities and larger cross-price elasticities for products that are similar in terms of their characteristics. The post-search taste shocks are, however, uncorrelated across products and merely introduce an additional source of randomness that (other things being equal) leads to prices and other characteristics impacting purchases relatively less. Hence, lower search costs lead to smaller cross-price elasticities. We also note that, while the sequential search model leads to elasticities that are similar to those from a full information probit model, search data can help with the estimation of heterogenous preference parameters as outlined in Section 4.3. Search data can therefore serve as a means to the end of estimating parameters that drive substitution patterns and elasticities.

In summary, elasticities generated from a sequential search model are similar in structure to those generated by a standard full information discrete choice model except for a set of product-specific uncorrelated error terms. These error terms attenuate the magnitude of own- and cross-price elasticities and the impact of these errors is larger when search costs are smaller.

5.3 Welfare

In a limited information setting, we have to take search costs into account when computing consumer welfare. We can write consumer surplus as

$$\mathbb{E}(CS) = \frac{1}{\alpha} \times \int \int \left(u_{ij}^{*} - H_i \times c\right) \times \phi(\boldsymbol{\varepsilon})\phi(\boldsymbol{\mu}) d\boldsymbol{\varepsilon} d\boldsymbol{\mu},$$

where u_{ij}^{*} denotes the utility of the chosen option and the second term denotes the total search costs the consumer incurs by searching H_i products from the set of products S_i.[18] Contrary to the equivalent expression in a perfect information context, u_{ij}^{*} does not necessarily denote the highest utility option in the market because the consumer might not discover the highest utility product in the presence of search frictions.

In principle, one could evaluate this expression by numerically integrating out the pre- and post-search taste shocks for a given vector of estimated parameters. This procedure requires the researcher to solve for the consumer's optimal search set and choice for each draw of parameters in order to compute both u_{ij}^{*} and S_i. A more convenient way to calculate welfare uses the effective value concept described above to derive an expression for consumer surplus:

$$\mathbb{E}(CS) = \frac{1}{\alpha} \times \int \int \max \left[X_j' \boldsymbol{\beta} - \alpha p_{ij} + min(\varepsilon_{ij}, g(c)) + \mu_{ij}\right] \times \phi(\boldsymbol{\varepsilon})\phi(\boldsymbol{\mu}) d\boldsymbol{\varepsilon} d\boldsymbol{\mu}. \tag{13}$$

To compute this expression, the reseacher is still required to simulate draws of the taste shocks, but it is not necessary to solve for consumers' optimal search and purchase behavior.

The welfare expression in equation (13) also highlights the similarity to the well-known surplus expression in the case of a perfect information discrete choice model. The only difference is the inclusion of the post-search error term ε_{ij}, whose influence on welfare is reduced in the presence of search costs. In particular, higher search costs imply a smaller value of $g(c)$ and hence the distribution of the post-search error term is truncated from above at a lower value. As search costs decrease, consumers search more and are more likely to discover post-search error terms. Thus, welfare increases by giving consumers access to additional utility from the post-search error realization.

5.4 Search Cost Counterfactuals

One class of counterfactuals that a researcher might be interested in conducting are counterfactuals that analyze the role of information on market outcomes and consumer welfare. One important counterfactual is one that quantifies the role of search frictions by analyzing behavior in the counterfactual setting in which search costs for all products are equal to zero. It is easy to simulate the changes in market shares resulting from a removal of search costs based on equation (12). Similarly, the researcher can compute welfare changes using the expression in equation (13). In both cases, the removal of search costs leads to a removal of the truncation of the post-search error and there-

[18] When search costs vary across products, the search cost component of the welfare expression becomes $\sum_{k \in S_i} c_j$.

fore $min(\varepsilon_{ij}, g(c))$ simply equals the post-search taste shock ε_{ij}. In settings with asymetric search costs across products (as discussed in Section 4.2), some products will benefit from the removal of search costs more than others. Welfare will unambiguously increase when search frictions are removed because consumers learn about all post-search errors rather than only a subset of them. In a similar vein, a researcher can analyze the consequence from a reduction in (rather than an elimination of) search costs using the same formulas.

A second class of counterfactuals are those that assess the role of characteristics that impact search costs. For example, search costs can be modeled as a function of advertising (Ursu et al. 2021a) or product rankings (Ursu 2018). This would then allow the researcher to analyze alternative product rankings or changes in advertising intensity across products. These types of changes alter search costs for different products by changing the characteristics that enter the search cost expression. Therefore, changes in market shares and welfare can still be analyzed using the formulas provided above.

Some of our earlier results are relevant when conducting this type of counterfactual. Being able to estimate the post-search taste shock variance is important for all counterfactuals outlined above because welfare effects depend on the estimated or normalized value of the post-search taste shock variance (see Section 3.3). More precisely, a change in search costs will alter the trunction point in the $min(\varepsilon_{ij}, g(c_{ij}))$ expression. A given reduction in search costs will have a different impact depending on the variance of the post-search error term distribution. The relevance of the error term variance is easiest to illustrate for the case of a welfare analysis when reducing search costs to zero. If the post-search variance is large, then the consumer will get access to an error term with heavier tails, which will lead to a larger welfare increase due to the max-operator in the welfare expression. That is, with a higher variance of post-search errors, welfare is larger because consumers are more likely to obtain a favorable draw for at least one product. Moreover, our identification results with regards to product-specific search costs are also relevant here. As outlined in Section 4.2, the researcher can identify the impact of characteristics on search costs and preferences separately. Hence, it is possible to allow a variable such as advertising to enter both utility and search costs. This distinction matters for assessing welfare changes due to changes in advertising: consumers benefit directly through the impact of advertising on utility and also change their search patterns through the influence of advertising on search costs.

6 Estimation

We present different estimation approaches in a general fashion without relying on a specific parameterization of utility and search costs, except for an additive-form utility with a pre- and a post-search taste shock as described in Section 2.3. We denote the parameters to be estimated by a vector $\boldsymbol{\theta}$. This parameter vector might comprise a price coefficient, preference weights on various product characteristics, and search costs in a homogenous model or the parameters governing the random coefficient distributions of various preference weights as well as search costs in a model that allows for unobserved heterogeneity.

6.1 Estimation Inequalities

Recall from Section 2.1 that the set of available options is denoted by $\Im = \{1, \ldots, J\}$ and the outside option is indicated by $j = 0$ with $u_{i0} = \mu_{i0}$, i.e., $\xi_{i0} = 0$ and the consumer knows the realized utility of the outside option prior to search.

For each consumer, we observe the following decisions: the set of searched products S_i, the order of searches and the identity of the purchased product y_i. Let $\bar{S}_i$ denote the set of products the consumer did not search. We order the options searched by consumer i by their observed order of search so that $h = 1$ corresponds to the product searched first and so on. The searched set is given by $S_i = \{1, ..., H_i\}$ where H_i denotes the total number of searches.[19]

The optimal decision rules described in Section 2.2.2 fully describe optimal search and purchase behavior. According to the *selection rule*, it must be that products are searched in decreasing order of reservation utilities:

$$z_{ih} \geq \max_{k \in \Im \setminus \{1,\ldots,h\}} z_{ik}, \qquad \forall h \in S_i. \tag{14}$$

In addition, the *stopping rule* imposes the following two restrictions: for the set of searched options, it must be that

$$z_{ih} \geq \max_{k=0}^{h-1} u_{ik}, \qquad \forall h \in S_i. \tag{15}$$

In contrast, for the options that were not searched, it must be that

$$\max_{h \in S_i \cup \{0\}} u_{ih} \geq \max_{l \in \bar{S}_i} z_{il}. \tag{16}$$

Finally, consistent with the *choice rule*, if the consumer chooses y_i, then her utility from this option is larger than that of any other searched product (including the outside option), i.e.,

$$u_{iy_i} \geq \max_{h \in S_i \cup \{0\}} u_{ih}. \tag{17}$$

When data are conditional on at least one search, the researcher typically assumes that the first search performed by a consumer is free (e.g., Honka 2014, Honka and Chintagunta 2017). If this additional assumption is imposed, the first stopping condition only applies to $h \in \{2, \ldots, H_i\}$. This assumption is not necessary when consumers who do not search are observed in the data (e.g., in Ursu 2018).

Next, we re-write the conditions governing search and purchase behavior, i.e., equations (14) to (17), in "differenced" form. This will be useful in the following sections.

[19] We note that each consumer may search products in a different order, so the product searched first, second, etc. by consumer i may not coincide with the order of searches made by another consumer.

$$
\begin{aligned}
\nu_{i,1h} &= z_{ih} - \max_{k \in \Im \setminus \{1,\dots,h\}} z_{ik} \qquad \forall h \in S_i, &(18)\\
\nu_{i,2h} &= z_{ih} - \max_{k=0}^{h-1} u_{ik} \qquad \forall h \in S_i, &(19)\\
\nu_{i,3} &= \max_{h \in S_i \cup \{0\}} u_{ih} - \max_{l \in \bar{S}_i} z_{il}, &(20)\\
\nu_{i,4} &= u_{iy_i} - \max_{h \in S_i \cup \{0\}} u_{ih}. &(21)
\end{aligned}
$$

If a condition is fulfilled, then $\nu \geq 0$ for that particular condition-consumer combination. In data sets in which consumers who do not search are also observed, there are H_i order and H_i continuation decisions for all $h \in S_i$ as well as one stopping and one purchase decision, i.e., a total of $2 \cdot H_i + 2$ conditions for consumer i. In data sets which are conditional on at least one search, there are H_i order and $H_i - 1$ continuation decisions for all $h \in S_i$ as well as one stopping and one purchase decision, i.e., a total of $2 \cdot H_i + 1$ conditions for consumer i.

The expressions above describe optimal consumer behavior based on reservation utilities and realized utilities, which, in turn, depend on the model parameters $\boldsymbol{\theta}$. The probability of observing a certain outcome in the data for consumer i is given by:

$$
\begin{aligned}
L_i(\boldsymbol{\theta}) = \; & Pr(\underbrace{z_{ih} \geq \max_{k \in \Im \setminus \{1,\dots,h\}} z_{ik} \; \forall h \in S_i}_{\text{selection rule}} \\
& \cap \underbrace{z_{ih} \geq \max_{k=0}^{h-1} u_{ik} \; \forall h \in S_i \; \cap \max_{h \in S_i \cup \{0\}} u_{ih} \geq \max_{l \in \bar{S}_i} z_{il}}_{\text{stopping rule}} \\
& \cap \underbrace{u_{iy_i} \geq \max_{h \in S_i \cup \{0\}} u_{ih}}_{\text{choice rule}}).
\end{aligned}
$$

The full likelihood function is obtained by summing over the consumer-specific loglikelihood contributions $\log L_i(\boldsymbol{\theta})$.

Estimation consists of maximizing the likelihood function. Since selection, stopping, and choice decisions are not made independently of each other, the likelihood function does not have a closed-form expression. Thus, in the following, we discuss several simulation-based estimation approaches. For all these approaches, to evaluate the likelihood for a given parameter vector guess, we need to take draws of the taste shocks (and potentially of preference parameters and search costs in a model with unobserved heterogeneity) and calculate the probability that the selection, stopping, and choice conditions are fulfilled. Moreover, to calculate the likelihood, we need to first derive reservation utilities to be able to evaluate the expressions that enter the likelihood function. In the next section, we first describe how reservation utilities can be calculated and then discuss methods for computing the likelihood function via simulation. Finally, we present results from a set of

Monte Carlo simulations that compare the different estimation approaches in terms of precision and computational speed in Appendix B. These simulations are based on Matlab codes for all the estimation methods described below and are publicly available.

6.2 Computing Reservation Utilities

Recall that reservation utilities z_{ij} are defined as follows:

$$\int_{z_{ij}}^{\infty} (u_{ij} - z_{ij}) dF_{ij}(u_{ij}) = c_{ij},$$

i.e., the reservation utility is the value that equates the marginal cost of searching with the expected marginal benefit of doing so. Going forward, we focus on the most common specification in which both μ_{ij} and ε_{ij} follow standard normal distributions, i.e., $\mu_{ij} \sim N(0,1)$ and $\varepsilon_{ij} \sim N(0,1)$. Under the assumption of standard normal post-search taste shocks, reservation utilities can be computed using the following equation

$$z_{ij} = \xi_{ij} + \mu_{ij} + m(c_{ij}) \tag{22}$$

where the value of $m(c_{ij})$ is obtained by solving the following equation (we drop the subscripts on the search cost term because the equation holds for any i and j):

$$c = \phi(m) + m \times \Phi(m) - m. \tag{23}$$

Weitzman (1979) showed that a unique solution for equation (23) exists. Thus, to compute reservation utilities, we can invert this relation, solve for m, and then compute the reservation utility using equation (22). Importantly, m only depends on search costs, but not on preference parameters, which enter reservation utilities linearly. Therefore, the only complication when calculating reservation utilities is the non-linear relationship between m and c which requires the researcher to solve for m for a given level of search costs. During estimation, one needs to solve for m repeatedly for every guess of the parameters governing search costs. Therefore, a fast method to establish the mapping between m and c is important in terms of lowering the computational burden of the estimation procedure.

A first approach, proposed by Kim et al. (2010) and used extensively in the literature (see, e.g., Chen and Yao 2017; Ursu 2018), addresses the computational burden associated with solving equation (23) by pre-computing the mapping between m and c and saving it in a look-up table. More specifically, the look-up table method involves the following steps. First, for values of m over some interval, evaluate equation (23) and solve for the value of search costs that corresponds to each value of m. Then, save this relation in a table with one column indicating the value of m and another column indicating the corresponding search cost level. Finally, when estimating the model, for a given search cost level evaluated by the optimization routine, look up the corresponding value of m and use it to calculate the reservation utility (using equation (22)). If search costs do not exactly match a value in the look-up table, then use linear interpolation between the two nearest grid-points to obtain the relevant level of m. Because the look-up table is based on a finite grid of search cost values, the method introduces error from using linear interpolations for search cost values that are not equal to grid-point values. The two alternative approaches that we discuss next both involve the use of a fast algorithm to solve for reservation utilities thereby avoiding the approximation error arising from the look-up table method.

A second approach proposed by Jiang et al. (2021) utilizes Newton's method to compute reservation utilities. They also start with equation (23). Solving for m for a given c is equivalent to finding the solution to the following function:

$$q(m) = (1 - \varPhi(m)) \left(\frac{\phi(m)}{1 - \varPhi(m)} - m \right) - c = 0$$

Newtons' method uses numerical analysis to find successively better approximations to the root of a function. The algorithm starts with an initial guess and iteratively finds the next guess as $m_{k+1} = m_k - \frac{q(m_k)}{q'(m_k)}$. After plugging in $q'(m)$ and rearranging terms, the next guess in the iteration can be simplified to

$$m_{k+1} = \frac{\phi(m_k) - c}{1 - \varPhi(m_k)}.$$

The iteration process stops when the difference between m_{k+1} and m_k falls below a threshold determined by the researcher. Jiang et al. (2021) use a threshold of e^{-10} in their empirical analysis and find that convergence is fast, requiring only a small number of iterations. Once a solution for m has been calculated, the reservation utility can be computed using equation (22).

A third approach proposed by Elberg et al. (2018) is to use a contraction mapping to solve for m using a re-arranged version of equation (23):

$$m = -c + \phi(m) + m \times \Phi(m).$$

The authors show that $\Gamma(m) = -c + \phi(m) + m \times \Phi(m)$ constitutes a contraction mapping. The paper also shows that this relationship is not unique to normally distributed errors and holds for a larger class of distributions. Moreover, Elberg et al. (2018) show that one can derive a closed-form solution for the reservation utility under the less commonly used assumption of a logistic distribution for the post-search error term.

We note that the second and third methods avoid any error from using linear interpolation; however, both approaches involve defining a convergence threshold and can result in numerical errors when the threshold is "too loose." In practice, both methods appear to converge quickly and allow researchers to set tight convergence thresholds thus avoiding numerical problems. In Appendix B, we report results from Monte Carlo simulations which show that all three methods lead to similar estimates and standard errors. We find that the most commonly used look-up table method is significantly faster than the other methods.

A final approach (see Morozov, 2022 and Greminger 2022) to dealing with the mapping from reservation utilities to search costs is to directly estimate $g(c)$ rather than the underlying search costs c. Because $g(c)$ enters the reservation utility expressions linearly, this approach avoids having to repeatedly solve for the non-linear relationship that translates search costs into $g(c)$. Instead, the researcher only applies the mapping once after the estimation by inverting the estimated $g(c)$ term into the implied value of search costs.

6.3 Estimation Approaches

Next, we present four approaches to estimate preference and search cost parameters in a sequential search model. We focus on approaches based on simulated maximum likelihood estimation (SMLE), namely the crude frequency simulator (e.g., Chen and Yao 2017), the kernel-smoothed frequency simulator (e.g., Honka 2014; Honka and Chintagunta 2017; Ursu 2018; Ursu et al. 2021c), the GHK method (e.g. Chung et al. 2019; Jiang et al. 2021), and importance sampling (e.g., Morozov et al. 2021).

6.3.1 Crude Frequency Simulator

The estimation procedure for the crude frequency simulator (also known as the accept-reject simulator) contains the steps below. We index realized utilities, reservation utilities, and other quantities that depend on parameter draws using a d superscript.

1. Take $d = \{1, \ldots, D\}$ sets of draws of μ_{ij} and ε_{ij} (each set of draws contains one draw of μ_{ij} and one draw of ε_{ij}) for each consumer-product combination, i.e., $D \times J \times N$ sets of draws.[20]

2. For a given guess of parameters $\boldsymbol{\theta}$, compute u_{ij}^d and z_{ij}^d for each set of draws d and each consumer-product combination.

3. Calculate the expressions in equations (18) to (21) for each set of draws d and each consumer. Compute the likelihood contribution for each consumer and draw:

$$L_i^d = \left[\prod_{h \in S_i} \mathbf{1}(\nu_{i,1h}^d > 0)\right] \times \left[\prod_{h \in S_i} \mathbf{1}(\nu_{i,2h}^d > 0)\right] \times \mathbf{1}(\nu_{i,3}^d > 0) \times \mathbf{1}(\nu_{i,4}^d > 0).$$

4. Compute $L_i = \frac{1}{D} \sum_{d=1}^{D} L_i^d$ for each consumer.

5. Compute $LogL = \sum_{i=1}^{N} \log(L_i)$.

To find the parameter values that maximize the likelihood, the researcher has to take draws of μ_{ij} and ε_{ij} (as well as draws of any random coefficients) in step 1 and then go through the remaining steps for an initial parameter guess. After evaluating the likelihood for the initial parameter guess,

[20] If the model includes random coefficients in preferences and/or search costs, take $D \times N$ draws for those coefficients as well.

the parameter guess is updated and steps 2 to 6 are repeated. The first step of generating draws does not have to be repeated during the estimation.[21] The second step involves calculating reservation utilities which we discussed in the previous subsection. These calculations have to be repeated for each parameter guess and therefore need to be relatively fast.

The crude frequency simulator approximates the consumer-specific likelihood contribution by the proportion of draws that satisfy the search and purchase conditions. This simulator is unbiased and straightforward to use. One example of a paper that has implemented this estimation approach in the context of a sequential search model is Chen and Yao (2017). The frequency estimator has two downsides. First, the frequency estimator can generate likelihood contributions L_i that are equal to zero if $L_i^d = 0$ for all draws. In this case, the log-likelihood function cannot be computed. Second, the simulated likelihood is not a smooth function in the parameters $\boldsymbol{\theta}$. Thus, the simulated likelihood is not differentiable and small changes in the parameters might not lead to any change in the simulated likelihood function. To avoid zero-valued likelihood contributions and to be able to use common optimization routines that rely on derivates, the researcher would need to take a large number of draws which will increase the computational burden of estimation (see the simulation results in Table B1 in Appendix B). The estimation routines that we present in the following subsections improve upon the crude frequency estimator by providing estimation approaches that avoid both issues.

6.3.2 Kernel-Smoothed Frequency Simulator

The kernel-smoothed frequency simulator (also known as the logit-smoothed accept-reject simulator) overcomes the disadvantage of the crude frequency simulator in that the resulting loglikelihood is a smooth function and likelihood contributions are larger than zero. As a result, common optimization routines can be applied for estimation.[22] The approach requires the researcher to choose a kernel and scaling parameter(s). The most commonly used kernel is a multivariate scaled logistic cdf (Gumbel 1961), which applied to the sequential search model leads to the following expression for the consumer-draw-specific likelihood contribution:

$$L_i^d = \frac{1}{1 + \sum_{k=1}^{2} \sum_{h \in S_i} \exp\left(-\rho_k \nu_{i,kh}^d\right) + \sum_{k=3}^{4} \exp\left(-\rho_k \nu_{i,k}^d\right)}.$$

where ρ_k is a scaling parameter for condition(s) ν_k (see equations 18 - 21) and $\boldsymbol{\rho}$ denotes the vector of scaling parameters. The estimation procedure using the logit-smoothed frequency simulator involves the same steps as the crude frequency estimator with the exception of step 3 where L_i^d is calculated based on the expression above.

The kernel-smoothed simulator replaces the step function of the crude frequency simulator with a smoothed sigmoid function. The degree of smoothing is governed by the scaling parameter vector

[21] In the presence of random coefficients, we assume that these come from known distributions, such a normal distribution, and therefore the researcher can take draws from a standard normal distribution in step 1 and then convert the draws when updating the parameters governing the distribution during the estimation.

[22] See Train (2009); Geweke and Keane (2001); Hajivassiliou et al. (1996) for a thorough discussion of the estimator.

$\boldsymbol{\rho}$. Suppose the scaling parameters are very large. If all conditions are fulfilled and hence $\nu_k > 0 \; \forall k$, then $L_i^d \rightarrow 1$. If one of the conditions is not fulfilled and hence $\nu_k < 0$ for some k then $L_i^d \rightarrow 0$. Therefore, for large values of the scaling parameters, the kernel-smoothed simulator approaches the crude frequency estimator. As a result, this simulator is asymptotically unbiased (see McFadden 1989, Train 2009). For smaller values of $\boldsymbol{\rho}$, the likelihood contribution does not equal zero even if some conditions take on a negative value and hence the estimator allows for deviations from the optimality conditions. However, finite scaling parameters lead to smoothness in the objective function and avoid zero-valued likelihood contributions.

The value of the scaling parameters ρ_k needs to be chosen by the researcher. As explained above, choosing larger values of ρ_k approximates the crude frequency simulator better, resulting in an unbiased estimate. However, values of ρ_k that are too large reintroduce the numerical problems common when optimizing a non-smooth function (such as in the case of the crude frequency simulator). There is little guidance provided in choosing these parameters, but as suggested by Train (2009), the best approach is for researchers to experiment with different values. Thus, prior work tests different values for the scaling parameters using Monte Carlo simulations, and then estimates the model with the scaling parameters that best recover the primitives of the model in simulated data (Honka 2014; Honka and Chintagunta 2017; Ursu 2018; Ursu et al. 2021c,a).

We also note that it is not necessary that each component of the likelihood function (i.e., the stopping, selection, and choice rules) have the same scaling parameter. Prior work (see Ursu et al. 2020) has experimented with different weights and shown that a different scaling parameter for every search rule and for the choice rule can achieve a better performance in certain empirical applications.

6.3.3 GHK

Another approach to estimating a sequential search model is the GHK simulator (named after Geweke 1989; Hajivassiliou and McFadden 1998; Keane 1994). The GHK simulator has the same two advantages as the smoothed frequency simulator over the crude frequency simulator: it does not produce zero-valued likelihood contributions and results in a smooth likelihood. Compared to the smoothed frequency simulator, the GHK simulator has the additional advantages that no smoothing parameters have to be chosen and that it is more efficient, i.e., fewer draws are needed. However, as described below and in Appendix A, it is more complex to implement the GHK simulator than the smoothed frequency simulator, e.g., the likelihood contributions are calculated differently depending on whether the consumer purchased the outside option, the last-searched product, or any other product.[23]

Chung et al. (2019) and Jiang et al. (2021) apply a GHK simulator in the context of a sequential search model. The following description of the GHK simulator roughly follows the discussion in Jiang et al. (2021), with the difference being that the utility of the outside option is given by

[23]A related approach that also relies on sequentially drawing truncated errors terms is the Bayesian estimation approach implemented in Morozov (2022).

$u_{i0} = \mu_{i0}$ instead of $u_{i0} = 0$. As before, we define products in order of search, with $h = \{1 \ldots H\}$ being the searched products (subscript i dropped for clarity) and $\bar{S}_i$ denoting the set of unsearched products. We also rely on the following two expressions for the utility function and the reservation utility throughout the derivations:

$$\begin{aligned} u_{ij} &= \xi_{ij} + \mu_{ij} + \varepsilon_{ij} \\ z_{ij} &= \xi_{ij} + \mu_{ij} + m(c). \end{aligned}$$

Here, we present the likelihood expression for the case of a consumer who purchased the outside option. The other cases are described in Appendix A.

1. Take D draws for μ_{i0} from its distribution with no bounds.[24]

2. For the last searched option H, take D draws for μ_{iH} from its distribution with a lower bound $\underline{b}_{iH} = \mu_{i0}^d - \xi_{iH} - m(c)$.

3. For all searched options but the last one, take D draws for $\mu_{i,H-1}, \ldots, \mu_{,i1}$ recursively with a lower bound $\underline{b}_{ih} = \xi_{i,h+1} + \mu_{i,h+1}^d - \xi_{ih}$.

4. Compute

$$\begin{aligned} L_i^d \quad = \quad & \underbrace{\prod_{l \in \bar{S}_i} \Phi\left(\mu_{i0}^d - \xi_{il} - m(c)\right)}_{stopping\ rule\ for\ unsearched\ options} \times \underbrace{\left(1 - \Phi\left(\mu_{i0}^d - \xi_{iH} - m(c)\right)\right)}_{continuation\ rule\ for\ last\ searched\ option} \\ & \times \underbrace{\prod_{h=1}^{H-1} \left(1 - \Phi\left(\xi_{i,h+1} + \mu_{i,h+1}^d - \xi_{ih}\right)\right)}_{selection\ rule\ for\ searched\ options} \times \underbrace{\prod_{h=1}^{H} \Phi\left(\mu_{i0}^d - \xi_{ih} - \mu_{ih}^d\right)}_{choice\ rule}. \end{aligned}$$

5. Compute $L_i = \frac{1}{D}\sum_{d=1}^{D} L_i^d$ for each consumer.

6. Compute $LogL = \sum_{i=1}^{N} \log(L_i)$.

The likelihood contribution calculated in step 4 consists of a series of conditional and unconditional probabilities that represent the probabilities that the conditions in equations (18) to (21) hold, which we previously denoted as $\{\nu_{i,1h}, \nu_{i,2h}, \nu_{i,3}, \nu_{i,4}\}$. Since we condition on the choice of the outside option, we can simplify the various optimality conditions by setting the maximum realized utility at every step of the search process equal to u_{i0} (because the outside option is always, i.e., at every step of the search process, available to consumers). Thus, for non-purchasers, the conditions $\{\nu_{i,1h}, \nu_{i,2h}, \nu_{i,3}, \nu_{i,4}\}$ simplify as follows:

[24]If the model includes random coefficients or heterogenous search costs, take draws of those parameters as well.

$$
\begin{aligned}
\nu_{i,1h} &= z_{ih} - \max_{k \in \Im \setminus \{1,\dots,h\}} z_{ik} > 0 \qquad \forall h \in S_i, \\
\nu_{i,2h} &= z_{ih} - u_{i0} > 0 \qquad \forall h \in S_i, \\
\nu_{i,3} &= u_{i0} - \max_{l \in \bar{S}_i} z_{il} > 0, \\
\nu_{i,4} &= u_{i0} - \max_{h \in S_i} u_{ih} > 0.
\end{aligned}
$$

We obtain the first term in L_i^d ("stopping rule for unsearched options") by evaluating the probability that condition $\nu_{i,3}$ holds (conditional on a set of draws), which states that the reservation utilities of all unsearched products are smaller than the utility of the outside option:

$$
\begin{aligned}
Pr(\max_{l \in \bar{S}_i} z_{il} < u_{i0}) &= Pr(\mu_{il} + \xi_{il} + m(c) < \mu_{i0}^d \ \ \forall l \in \bar{S}_i) \\
&= Pr(\mu_{il} < \mu_{i0}^d - \xi_{il} - m(c) \ \ \forall l \in \bar{S}_i) \\
&= \prod_{l \in \bar{S}_i} \Phi\left(\mu_{i0}^d - \xi_{il} - m(c)\right)
\end{aligned}
$$

where the last line follows from the fact that μ_{il} is independently and normally distributed.

The second term in L_i^d ("continuation rule for last searched option") is equal to the probability that the consumer searches the final product H in her search sequence. This probability corresponds to condition $\nu_{i,2H}$ above, the last continuation decision in the search sequence.

$$
\begin{aligned}
Pr(z_{iH} > u_{i0}) &= Pr(\mu_{iH} + \xi_{iH} + m(c) > \mu_{i0}^d) \\
&= Pr(\mu_{iH} > \mu_{i0}^d - \xi_{iH} - m(c)) \\
&= 1 - Pr(\mu_{iH} < \mu_{i0}^d - \xi_{iH} - m(c)) \\
&= 1 - \Phi\left(\mu_{i0}^d - \xi_{iH} - m(c)\right)
\end{aligned}
$$

The third term ("selection rule for searched options") corresponds to the order conditions for products 1 to $H-1$, conditional on the the continuation decision for product H and the order conditions for products later in the search sequence. Taking the case of product $H-1$ as an example, the conditional probability is equal to:

$$
\begin{aligned}
Pr(z_{iH-1} > z_{iH} \mid \mu_{iH} > \mu_{i0}^d - \xi_{iH} - m(c)) &= Pr(\mu_{iH-1} + \xi_{iH-1} > \xi_{iH} + \mu_{iH}^d) \\
&= Pr(\mu_{iH-1} > \xi_{iH} + \mu_{iH}^d - \xi_{iH-1}) \\
&= 1 - Pr(\mu_{iH-1} < \xi_{iH} + \mu_{iH}^d - \xi_{iH-1}) \\
&= 1 - \Phi\left(\xi_{iH} + \mu_{iH}^d - \xi_{iH-1}\right)
\end{aligned}
$$

Importantly, the conditioning statement on the left-hand side is fulfilled because the draw μ_{iH}^d comes form an appropriately truncated distribution (see step 2).

Probabilities for the other order conditions are calculated in a similar fashion. For example, to evaluate $Pr(z_{iH-2} > z_{iH-1})$, we need to condition on $(z_{iH-1} > z_{iH})$ which is done by drawing μ_{iH-1}^d as described in step 3. Working backwards through the products in the search sequence, the third term describes the probability that the order conditions $\nu_{i,1h}$ for products 1 to $H-1$ hold. The continuation conditions $\nu_{i,2h}$ hold because we established that $z_H > u_{i0}$ and $z_h > z_H \, \forall h \in S_i$ due to the order condition. It follows that $z_h > u_{i0} \forall h \in S_i$. Finally, the order condition for product H holds because we established that $z_H > u_{i0}$ and $\max_{l \in \bar{S}_i} z_{il} < u_{i0}$ and therefore it follows that $z_H > \max_{l \in \bar{S}_i} z_{il}$.

The only condition not captured so far is the purchase condition $\nu_{i,4}$ which represents the final term in the likelihood contribution L_i^d ("choice rule"):

$$\begin{aligned} Pr(u_{i0} > \max_{k \in S_i} u_{ik} \mid ...) &= Pr(\mu_{ih}^d + \xi_{ih} + \varepsilon_{ih} < \mu_{i0} \; \forall \; h \in S_i) \\ &= Pr(\varepsilon_{ih} < \mu_{i0} - \xi_{ih} - \mu_{ih}^d \; \forall \; h \in S_i) \\ &= \prod_{h=1}^{H} \Phi\left(\mu_{i0} - \xi_{ih} - \mu_{ih}^d\right) \end{aligned}$$

where the conditioning (not written out for brevity) is on all the other conditions described above that lead to the truncated draws of μ_{ih} for all searched products.

Taken together, the expressions above fulfill all search and choice conditions $\{\nu_{i,1h}, \nu_{i,2h}, \nu_{i,3}, \nu_{i,4}\}$ and jointly yield the likelihood contribution for a consumer who chooses the outside option. We describe the analogous procedure for consumers who purchased one of the available products in Appendix A.

6.3.4 Importance Sampling

Importance sampling is an alternative estimation approach and beneficial in a model that allows for unobserved heterogeneity via random coefficients in all model parameters, i.e., both preference and search cost parameters. The main idea behind importance sampling in the context of search models is that it allows the researcher to compute reservation utilities for a larger number of simulation draws only once at the beginning of the estimation routine. Similar to applications of importance sampling to dynamic demand models (e.g., Hartmann, 2006), we are thus able to pre-compute a complicated object (reservation utilities in this case), rather than having to repeatedly calculate it during estimation. Our discussion closely follows the application of importance sampling to search models in Morozov et al. (2021). For a more detailed discussion of importance sampling in demand estimation and related applications, we refer the reader to Ackerberg (2009).

We modify our notation slightly to focus on the case of unobserved heterogeneity in all parameters. We let $\boldsymbol{\theta_i}$ denote the vector of consumer-specific preference and search cost parameters and we let $f(\boldsymbol{\theta}|\boldsymbol{\Omega})$ denote the distribution from which these parameters are drawn. The vector $\boldsymbol{\Omega}$ denotes the parameters that govern these distributions (such as the mean and variance of normally distributed random coefficients). In the estimation, we aim to find the values of $\boldsymbol{\Omega}$ that maximize the following likelihood function

$$\begin{aligned} & \prod_{i=1}^{N} \int \left(L_i(\boldsymbol{\theta_i}) \times f(\boldsymbol{\theta_i}|\boldsymbol{\Omega})\right) d\boldsymbol{\theta_i} \\ = & \prod_{i=1}^{N} \int \left(L_i(\boldsymbol{\theta_i}) \times \frac{f(\boldsymbol{\theta_i}|\boldsymbol{\Omega})}{\tilde{f}(\boldsymbol{\theta_i})}\right) \tilde{f}(\boldsymbol{\theta_i}) d\boldsymbol{\theta_i}. \end{aligned}$$

Going from line 1 to line 2 in the above equation, we multiply and divide the consumer-specific likelihood contribution L_i by the density function $\tilde{f}(\boldsymbol{\theta_i})$. We choose the function $\tilde{f}(\boldsymbol{\theta_i})$, the "proposal density" such that it does not depend on the parameter vector $\mathbf{\Omega}$ and has non-zero density over the support of $\boldsymbol{\theta}$ (see Ackerberg (2009)). Then, we take draws of preference and search cost parameters based on $\tilde{f}(\boldsymbol{\theta_i})$ and calculate the likelihood function via simulation:

$$L(\boldsymbol{\Omega}) = \prod_{i=1}^{N} \frac{1}{M} \sum_{m=1}^{M} \left(L_i(\boldsymbol{\theta_i^m}) \cdot \frac{f(\boldsymbol{\theta_i^m}|\boldsymbol{\Omega})}{\tilde{f}(\boldsymbol{\theta_i})}\right)$$

where $\boldsymbol{\theta_i^m}$ denotes the m-th draw of parameters for consumer i, and $L_i(\boldsymbol{\theta_i^m})$ denotes the simulated consumer-specific likelihood contribution:

$$L_i(\boldsymbol{\theta_i^m}) = \frac{1}{D} \sum_{d=1}^{D} L_i^d(\boldsymbol{\theta_i^m}, \boldsymbol{\mu_i^d}, \boldsymbol{\varepsilon_i^d})$$

where $\boldsymbol{\mu_i^d}$ and $\boldsymbol{\varepsilon_i^d}$ are vectors of taste shock draws. We can then estimate parameters $\boldsymbol{\Omega}$ by maximizing the simulated likelihood $L(\boldsymbol{\Omega})$.

To re-cap, the estimators involves the following steps:

1. For each consumer take M draws of preference / search cost parameters $\boldsymbol{\theta_i^m}$. For each consumer / parameter draw, take D draws of taste shocks $(\boldsymbol{\mu_i^d}, \boldsymbol{\varepsilon_i^d})$.

2. Compute u_{ij}^d and z_{ij}^d for each set of draws d and each consumer-product combination.

3. Calculate $\nu_{i,1h}^d, \nu_{i,2h}^d, \nu_{i,3}^d, \nu_{i,4}^d$.

4. Calculate the likelihood contribution for each draw:

$$L_i^d = \left[\prod_{h \in S_i} \mathbf{1}(\nu_{ih,1}^d > 0)\right] \times \left[\prod_{h \in S_i} \mathbf{1}(\nu_{ih,2}^d > 0)\right] \times \mathbf{1}(\nu_{i,3}^d > 0) \times \mathbf{1}(\nu_{i,4}^d > 0).$$

5. Compute $L_i(\boldsymbol{\theta_i^m}) = \frac{1}{D} \sum_{d=1}^{D} L_i^d$ for each consumer.

6. Compute $L(\boldsymbol{\Omega}) = \prod_{i=1}^{N} \frac{1}{M} \sum_{m=1}^{M} \left(L_i(\boldsymbol{\theta_i^m}) \cdot \frac{f(\boldsymbol{\theta_i^m}|\boldsymbol{\Omega})}{\tilde{f}(\boldsymbol{\theta_i})}\right)$.

The steps outlined above mirror those of the other estimators we described, except for one key difference. When applying a frequency estimator, one needs to iterate through steps 2 to 6 for every guess of parameters when trying to find the parameters that maximize the likelihood function. This involves re-calculating reservation utilities in each iteration of the estimation procedure. If the set of draws being used is large, the number of consumer / draw combinations and hence reservation utility calculations can be large, thus increasing the computational burden. When applying an importance sampling approach, steps 1 to 5 only have to be implemented once and hence reservation utilities also have to be calculated only one single time. This simplification is possible because the likelihood only depends on the parameter vector $\mathbf{\Omega}$ through the weights $f(\boldsymbol{\theta}_i^m|\boldsymbol{\Omega})/\tilde{f}(\boldsymbol{\theta}_i^m)$, which are used when aggregating likelihood contributions at each draw into the full likelihood expression in step 6. In practice, the importance sampling procedure allows the researcher to take a large number of draws. Morozov et al. (2021) use an importance sampling estimator with $D = 1,000$ and $M = 1,000$, which results in 100,000 simulation draws per consumer. Similar to the kernel-smoothed frequency estimator, the importance sampling method results in an objective function that is smooth in parameters because the weights $f(\boldsymbol{\theta}_i^m|\boldsymbol{\Omega})/\tilde{f}(\boldsymbol{\theta}_i^m)$ are continuous and differentiable in $\mathbf{\Omega}$.

It is in principle possible that some consumer-specific likelihood contributions L_i^d are equal to zero. The importance sampling procedure does not directly solve this particular issue that also plagued the crude frequency estimator. However, because likelihood contributions are computed only once at the beginning of the estimation routine, it is easy to increase the number of draws sufficiently to avoid any occurrence of zero-valued likelihood contributions.

When implementing the importance sampling approach above, the researcher needs to choose a proposal density, which can affect the variance of the estimator (see the discussion in Ackerberg 2009). One natural choice would be to set the proposal density equal to the true density function evaluated at some initial guess $\boldsymbol{\Omega^{init}}$, i.e. $\tilde{f}(\boldsymbol{\theta_i}) = f(\boldsymbol{\theta}_i^m|\boldsymbol{\Omega^{init}})$. Morozov et al. (2021) use such an approach to choose the proposal density, where $\boldsymbol{\Omega^{init}}$ is choosen such that some key moments, such as purchase and search probabilities, match their empirical counterparts.

The main downside of the importance sampling procedure is that its computational benefits can only be harnessed if the model allows for random coefficients in all preference and search cost parameters. Unobserved heterogeneity in parameters implies that we can take draws of all parameters once and then re-weight the draws during the estimation. Due to unobserved heterogeneity, all draws have a positive probability of occurring and the probability is shifted as a function of the parameters that govern the distribution of random coefficients. The possibility of re-weighting draws does not exist for parameters without random coefficients. While it is possible to use importance sampling in the context of a model that contains random coefficients only for a subset of coefficients, the procedure loses its computational advantage because the pre-calculation of likelihood contributions is not available anymore

7 Conclusion

The goal of this paper was to consolidate knowledge on the assumptions, properties, identification, and estimation of sequential search models à la Weitzman (1979) and to provide a unified treatment of these various aspects of the most frequently used framework for empirical research on consumer search. It is our hope that applied researchers can use this paper as a guide with regards to choosing the appropriate estimation method, understanding which parameterization of utility and search costs can be identified, and which counterfactuals can be assessed with a particular model specification.

References

Abaluck, J., G. Compiani, and F. Zhang (2022): "A Method to Estimate Discrete Choice Models that is Robust to Consumer Search," Working Paper.

Ackerberg, D. A. (2009): "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation," *Quantitative Marketing and Economics*, 7, 343–376.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 132, 1553–1592.

Armstrong, M. (2017): "Ordered Consumer Search," *Journal of the European Economic Association*, 15, 898–1024.

Berry, S., and P. Haile (2014): "Identification of Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

Berry, S., J. Levinsohn, and A. Pakes (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112, 68–105.

Branco, F., M. Sun, and J. M. Villas-Boas (2012): "Optimal Search for Product Information," *Management Science*, 58, 2037–2056.

——— (2016): "Too Much Information? Information Provision and Search Costs," *Marketing Science*, 35, 605–618.

Breen, R., K. Karlson, and A. Holm (2018): "Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models," *Annual Review of Sociology*, 44, 39–54.

Cameron, S., and J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106, 262–333.

Chen, Y., and S. Yao (2017): "Sequential Search with Refinement: Model and Application with Click-stream Data," *Management Science*, 63, 4345–4365.

Chick, S., and P. Frazier (2012): "Sequential Sampling with Economics of Selection Procedures," *Management Science*, 58, 1–16.

Ching, A., T. Hossain, S. Tehrani, and C. Zhao (2020): "How Do People Update Beliefs? Evidence from the Laboratory," Working Paper.

Choi, H., and C. Mela (2019): "Monetizing Online Marketplaces," *Marketing Science*, 38, 948–972.

Choi, M., A. Y. Dai, and K. Kim (2018): "Consumer Search and Price Competition," *Econometrica*, 86, 1257–1281.

Chung, J. H., P. Chintagunta, and S. Misra (2019): "Estimation of Sequential Search Model," Working Paper.

Compiani, G., G. Lewis, S. Peng, and W. Wang (2022): "Online Search and Product Rankings: A Double Index Approach," Working Paper.

De Los Santos, B. I., A. Hortacsu, and M. Wildenbeest (2012): "Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102, 2955–2980.

Dukes, A., and X. Liu (2015): "Online Shopping Intermediaries: The Strategic Design of Search Environments," *Management Science*, 62, 1064–1077.

Elberg, A., P. M. Gardete, R. Macera, and C. Noton (2018): "Dynamic Effects of Price Promotions: Field Evidence, Consumer Search, and Supply-Side Implications," *Quantitative Marketing and Economics*, 17, 1–58.

Gardete, P., and M. Hunter (2020): "Guiding Consumers through Lemons and Peaches: An Analysis of the Effects of Search Design Activities," Working Paper.

Geweke, J. (1989): "Bayesian Inference in Econometrics Models using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Geweke, J., and M. Keane (2001): "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics* ed. by Heckman, J., and Leamer, E. Volume 5: Elsevier, 1st edition, Chap. 56, 3463–3568.

Greminger, R. (2022): "Heterogeneous Position Effects and the Power of Rankings," Working Paper.

Gumbel, E. (1961): "Bivariate Logistic Distributions," *Journal of the American Statistical Association*, 56, 335–349.

Hajivassiliou, V., and D. McFadden (1998): "The Method of Simulated Scores for the Estimation of LDV Models," *Econometrica*, 66, 863–896.

Hajivassiliou, V., D. McFadden, and P. Ruud (1996): "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results," *Journal of Econometrics*, 72, 85–134.

Hartmann, W. R. (2006): "Intertemporal Effects of Consumption and their Implications for Demand Elasticity Estimates," *Quantitative Marketing and Economics*, 4, 325–349.

Häubl, G., B. Dellaert, and B. Donkers (2010): "Tunnel Vision: Local Behavioral Influences on Consumer Decisions in Product Search," *Marketing Science*, 29, 438–455.

Hodgson, C., and G. Lewis (2020): "You Can Lead a Horse to Water: Spatial Learning and Path Dependence in Consumer Search."

Honka, E., A. Hortascu, and M. Wildenbeest (2019): "Empirical Search and Consideration Sets," in *Handbook on the Economics of Marketing* ed. by Dube, J., and Rossi, P.: North Holland.

Honka, E. (2014): "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry," *The RAND Journal of Economics*, 45, 847–884.

Honka, E., and P. K. Chintagunta (2017): "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry," *Marketing Science*, 36, 21–42.

Hsiao, C. (1983): "Identification," in *Handbook of Econometrics, Volume I* ed. by Griliches, Z., and Intriligator, M.: North Holland.

JIANG, Z., T. CHEN, H. CHE, AND Y. WANG (2021): "Consumer Search and Purchase: An Empirical Investigation of Retargeting Based on Consumer Online Behaviors," *Marketing Science*, 40, 219–240.

JINDAL, P., AND A. ARIBARG (2021): "The Importance of Price Beliefs in Consumer Search," *Journal of Marketing Research*, 58, 321–342.

KE, T. T., Z.-J. M. SHEN, AND J. M. VILLAS-BOAS (2016): "Search for Information on Multiple Products," *Management Science*, 62, 3576–3603.

KE, T. T., AND J. M. VILLAS-BOAS (2019): "Optimal Learning Before Choice," *Journal of Economic Theory*, 180, 383–437.

KEANE, M. (1994): "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.

KIM, J. B., P. ALBUQUERQUE, AND B. J. BRONNENBERG (2010): "Online Demand under Limited Consumer Search," *Marketing Science*, 29, 1001–1023.

KOULAYEV, S. (2013): "Search With Dirichlet Priors: Estimation and Implications for Consumer Demand," *Journal of Business & Economic Statistics*, 31, 226–239.

MADDALA, G. (1983): in *Limited-Dependent and Qualitative Variables in Econometrics*: Cambridge, UK: Cambridge Univ. Press.

MATSUMOTO, B., AND F. SPENCE (2016): "Price Beliefs and Experience: Do Consumers' Beliefs Converge to Empirical Distributions with Repeated Purchases?," *Journal of Economic Behavior & Organization*, 126, 243–254.

MATZKIN, R. (2013): "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 5, 457–486.

MCCALL, J. (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics*, 84, 113–126.

MCFADDEN, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 995–1026.

MORAGA-GONZÁLEZ, J. L., Z. SÁNDOR, AND M. R. WILDENBEEST (2022): "Consumer Search and Prices in the Automobile Market," *Review of Economic Studies (forthcoming)*.

MOROZOV, I. (2022): "Measuring Benefits from New Products in Markets with Information Frictions," *Management Science (forthcoming)*.

MOROZOV, I., S. SEILER, X. DONG, AND L. HOU (2021): "Estimation of Preference Heterogeneity in Markets with Costly Search," *Marketing Science*, 40, 871–899.

DE LOS SANTOS, B., A. HORTAÇSU, AND M. R. WILDENBEEST (2017): "Search with Learning for Differentiated Products: Evidence from E-Commerce," *Journal of Business & Economic Statistics*, 35, 626–641.

SEILER, S. (2013): "The Impact of Search Costs on Consumer Behavior: A Dynamic Approach," *Quantitative Marketing and Economics*, 11, 155–203.

Stigler, G. (1961): "The Economics of Information," *Journal of Political Economy*, 69, 213–225.

Train, K. E. (2009): *Discrete Choice Methods with Simulation*: Cambridge university press.

Ursu, R., Q. Wang, and P. Chintagunta (2020): "Search Duration," *Marketing Science*, 39, 849–871.

Ursu, R. M. (2018): "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," *Marketing Science*, 37, 530–552.

Ursu, R., A. Simonov, and E. An (2021a): "Online Advertising as Passive Search," Working Paper.

Ursu, R., Q. Zhang, and E. Honka (2021c): "Search Gaps and Consumer Fatigue," Working Paper.

Ursu, R., Q. P. Zhang, and T. Erdem (2021b): "Prior Information and Consumer Search: Evidence from Eye-tracking," *Working Paper*.

Weitzman, M. L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, 47, 641–654.

Yao, S., W. Wang, and Y. Chen (2017): "TV Channel Search and Commercial Breaks," *Journal of Marketing Research*, 54, 671–686.

Yavorsky, D., E. Honka, and K. Chen (2021): "Consumer Search in the U.S. Auto Industry: The Value of Dealership Visits," *Quantitative Marketing and Economics*, 19, 1–52.

Appendix

A Additional Details on GHK Procedure

In the main part of the paper, we describe how to construct the consumer-specific likelihood contribution for consumers who choose the outside option. The procedure for consumer who purchase one of the available products is slightly different and outlined below. We describe the procedure separately for consumers who purchased the last product they searched and consumers who purchased a product that was searched earlier in the search sequence.

A.1 Likelihood for Consumers who Purchased the Product Searched Last

Starting with consumers who bought the option searched last, we can compute the likelihood contribution as follows:

1. Take D draws for μ_{il} and μ_{i0} from their distributions with no bounds
2. For the last searched option H, take D draws for μ_{iH} from its distribution with a lower bound $\underline{b}_{iH} = \max\left(\max_l\left(\xi_{il} + \mu_{il}^d\right) - \xi_{iH}\ ,\ \mu_{i0}^d - \xi_{iH} - m\left(c\right)\right)$ (truncation comes from order and part of the continuation conditions)
3. For all searched options but the last one, take D draws for $\mu_{i,H-1}, \ldots \mu_{,i1}$ recursively with a lower bound $\underline{b}_{ih}^{\mu} = \xi_{i,h+1} + \mu_{i,h+1}^d - \xi_{ih}$ (truncation comes from order condition)
4. For all searched options but the last one, take D draws for $\varepsilon_{i1}, \ldots, \varepsilon_{i,H-1}$ from the distribution with an upper bound $\bar{b}_{ih}^{\varepsilon} = \xi_{iH} + \mu_{iH}^d + m\left(c\right) - \xi_{ih} - \mu_{ih}^d$ (truncation comes from the continuation condition)
5. For each consumer, compute

$$
\begin{aligned}
L_i \quad = \quad & \frac{1}{D}\sum_{d=1}^{D} \underbrace{\left(1 - \Phi\left(\max\left(\max_l\left(\xi_{il} + \mu_{il}^d\right) - \xi_{iH}, \mu_{i0}^d - \xi_{iH} - m\left(c\right)\right)\right)\right)}_{selection\ rule\ for\ unsearched\ options\ and\ part\ of\ continuation\ rule\ for\ H} \\
& \times \underbrace{\prod_{h=1}^{H-1}\left(1 - \Phi\left(\xi_{i,h+1} + \mu_{i,h+1}^d - \xi_{ih}\right)\right)}_{selection\ rule\ for\ 1...H-1\ searched\ options} \\
& \times \underbrace{\prod_{h=1}^{H-1}\Phi\left(\xi_{iH} + \mu_{iH}^d + m\left(c\right) - \xi_{ih} - \mu_{ih}^d\right)}_{continuation\ rule\ for\ searched\ options} \\
& \times \underbrace{\left(1 - \mathit{\Phi}\left[\max\left(\mu_{i0}^d, \max_l z_{il}^d, \max_{h=1,\ldots,H-1} u_{ih}^d\right) - \xi_{iH} - \mu_{iH}^d\right]\right)}_{stopping\ and\ choice\ rules}
\end{aligned}
$$

A.2 Likelihood for Consumers who Purchased a Product Not Searched Last

For purchasers who bought an option that was not searched last, the likelihood contribution is calculated as follows:

1. Take D draws for μ_{il} and μ_{i0} from its distribution with no bounds

2. For the last searched option H, take D draws for μ_{iH} from its distribution with a lower bound $\underline{b}_{iH} = \max\left[\max_l\left(\xi_{il} + \mu_{il}^d\right) - \xi_{iH}, \mu_{i0}^d - \xi_{iH} - m\left(c\right)\right]$ (truncation comes from order and part of continuation condition)

3. For all searched options but the last one, take D draws for $\mu_{i,H-1}, \ldots \mu_{,i1}$ recursively with a lower bound $\underline{b}_{ih}^{\mu} = \xi_{i,h+1} + \mu_{i,h+1}^d - \xi_{ih}$ (truncation comes from order condition)

4. For the purchased option h^*, take D draws for ε_{ij} from its distribution with an upper bound $\bar{b}_{ih^*}^{\varepsilon} = z_{iH}^d - \xi_{ih*} - \mu_{ih*}^d$ and a lower bound $\underline{b}_{ih^*}^{\varepsilon} = \max\left(\max_l z_{il}^d, \mu_{i0}^d\right) - \xi_{ih*} - \mu_{ih*}^d$ (truncations come from continuation and part of the choice conditions)

5. For each consumer, compute

$$
\begin{aligned}
L_i \quad = \quad & \frac{1}{D}\sum_{d=1}^{D} \underbrace{\left(1 - \Phi\left(\max\left(\max_l\left(\xi_{il} + \mu_{il}^d\right) - \xi_{iH}, \mu_{i0}^d - \xi_{iH} - m\left(c\right)\right)\right)\right)}_{selection\ rule\ for\ unsearched\ options\ and\ part\ of\ continuation\ rule\ for\ H} \\
& \times \underbrace{\prod_{h=1}^{H-1}\left(1 - \Phi\left(\xi_{i,h+1} + \mu_{i,h+1}^d - \xi_{ih}\right)\right)}_{selection\ rule\ for\ 1 \ldots H-1\ searched\ options} \\
& \times \underbrace{\left(\Phi\left(z_{iH}^d - \xi_{ih*} - \mu_{ih*}^d\right) - \Phi\left(\max\left(\max_l z_{il}^d, \mu_{i0}^d\right) - \xi_{ih*} - \mu_{ih*}^d\right)\right)}_{continuation\ rule\ and\ part\ of\ choice\ rule\ for\ purchased\ option} \\
& \times \underbrace{\prod_{h\in\{1,\ldots,H\}\setminus h*} \Phi\left(u_{ih*}^d - \xi_{ih} - \mu_{ih}^d\right)}_{choice\ rule}
\end{aligned}
$$

B Monte Carlo Simulations

In this section, we report the results from a set of Monte Carlo simulations that compare the four estimation approaches (crude and kernel-smoothed frequency simulator, GHK, and importance sampling) and the three methods to compute reservation utilities (look-up table, Newton's method, and contraction mapping) in terms of their parameter recovery and computational time. For simplicity, we only report results for the different methods to compute reservation utilities for the kernel-smoothed frequency estimator, the most frequently used method in empirical work.[25] The code for generating the data and the various estimation procedures is publicly available.[26]

We generate data for 1,000 consumers who make search and purchase decisions in a setting with four brands and an outside option (with mean utility normalized to zero). To keep the model simple, the utility function only consists of brand intercepts. For the crude frequency simulator, the kernel-smoothed frequency simulator, and GHK, we generate data and estimate parameters based on homogenous preferences and search costs, i.e., we estimate four brand intercepts and a search cost parameter. For the importance sampling estimator, we generate data and estimate parameters based on heterogenous preferences and search costs, i.e., we assume that preferences and search costs are normally distributed and estimate the means and standard deviations of the random coefficients (assuming a diagonal variance-covariance matrix). As described in Section 6.3.4, the computational improvements of importance sampling only apply to models with heterogeneity in all model parameters (i.e., preferences and search costs). The computational time for importance sampling is therefore not directly comparable to the computational time for the three other procedures.

To estimate the model, we follow the steps described in Sections 6.2 and 6.3. Because of the non-smoothness of the likelihood function when using the crude frequency simulator, we report results when taking $D = 100$ and $D = 1,000,000$ draws of the error terms. We use $D = 100$ draws for the remaining three approaches. Further, for importance sampling, we take $M = 100$ draws of the preference and search cost parameters and $D = 100$ for each draw of preference parameters (i.e., a total of 10,000 draws per consumer). For the kernel-smoothed frequency simulator, we use the scaling vector $\rho_k = [-18, -4, -4, -7]$.[27] We repeat the estimation for each approach on 50 different data sets generated using the same true parameters, but different draws of the errors terms. All estimations start from a vector of zeros (for importance sampling, starting values for the standard deviations are set equal to one).

The results for the Monte Carlo simulations are displayed in Table B1. In column (1), we present the true parameters used to simulate the data; in columns (2) through (8), we show re-

[25] As we show below, the way in which reservation utilities are calculated has relatively little impact on parameter recovery. Moreover, we do not expect the reservation utility calculation to interact with the estimation method in an important way.

[26] Link: https://drive.google.com/file/d/1-GUVl-FtF_nkKQx-A6wjG0jQNwt3fqjm/view?usp=share_link.

[27] To find the scaling vector, we evaluated more than 50 different vectors and chose the values that minimized the distance between the true parameter vector and its estimate (in terms of MSE). Evaluating additional vectors may result in more precise parameter estimates.

sults for each of the estimation approaches. In parentheses, we report the standard deviations of these estimated coefficients across the 50 data simulations and estimations. Overall, all estimation approaches are able to recover the true parameters well, expect for the crude frequency estimator with a small number of draws. When using a larger number of draws (see column (3)), the crude frequency estimator is also able to recover parameters well. The standard deviations of the parameter estimates are slightly smaller for GHK relative to the crude and kernel-smoothed frequency estimators. The use of different methods for computing reservation utilities has little impact on the recovery of model parameters (see columns (4) to (6)).[28] In terms of computational time, we find that the look-up table method is significantly faster than other methods for calculating reservation utilities. GHK takes longer to run relative to a kernel-smoothed frequency estimator (when the look-up method is used).[29] The crude frequency estimator based on 1,000,000 error draws takes several orders of magnitude longer to run (close to one week) relative to the other approaches.

Finally, the importance sampling results show that parameters are well-recovered even though the approach is applied to a more complicated model with heterogeneity in all model parameters. The distance between the true and the estimated parameter vectors is smaller than for the other methods and the standard deviations of the mean utility and search cost parameters are also smaller. Importance sampling is computationally slower than most other methods, but uses a much larger number of draws than GHK and the kernel-smoothed frequency estimator. We re-iterate that the importance sampling results are hard to directly compare to the other methods because the underlying data-generating process is different. Morozov et al. (2021) compare estimation via importance sampling with a kernel-smoothed frequency estimator using simulated data from a model with heterogeneity in all parameters in the appendix of their paper. They find that the kernel-smoothed frequency estimator is significantly slower when using a modest number of draws and becomes prohibitively slow for large numbers of draws.

In summary, we conclude that the method for computing reservation utilities has relatively little impact on parameter recovery and that the look-up table method (implemented on a fine grid) is significantly faster than alternative methods.[30] A crude frequency estimator requires a much larger number of draws and therefore more computational time to recover parameters compared to other approaches. GHK recovers parameters better than a kernel-smoothed frequency estimator, but at the cost of a moderately longer run time. For models with heterogeneity, the importance sampling estimator recovers parameters well with a moderately higher run time than the other estimators (which are applied to a simpler model).

[28]The grid fineness for the look-up table method is 0.001, while the tolerance level for Newton's method and contraction mapping equals 10e-10.

[29]We note that the precision of the kernel-smoothed frequency estimator depends on the scaling vector and determining this scaling vector may in some cases take a significant amounts of time.

[30]Our results are based on estimation in Matlab and may not extend to other programming languages, such as R or Python.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimation Method		Crude Freq.	Crude Freq.	Kernel-Smooth.	Kernel-Smooth.	Kernel-Smooth.	GHK	Imp. Samp.
Reservation Value Calculation		Look-up Table	Look-up Table	Look-up Table	Newton	Contraction Mapping	Look-up Table	Look-up Table
	True Values	Estimates (Std. Dev.)	Estimates (Std. Dev.)	Estimates (Std. Dev.)	Estimates (Std. Dev.)	Estimates (Std. Dev.)	Estimates (Std. Dev.)	Estimates (Std. Dev.)
Preferences: Mean Utility								
Brand 1	1	0.67	0.97	0.87	0.87	0.87	0.93	0.99
		(0.36)	(0.11)	(0.10)	(0.10)	(0.10)	(0.08)	(0.03)
Brand 2	0.7	0.41	0.66	0.68	0.68	0.68	0.66	0.70
		(0.37)	(0.12)	(0.14)	(0.13)	(0.13)	(0.08)	(0.01)
Brand 3	0.5	0.29	0.47	0.51	0.51	0.51	0.45	0.50
		(0.36)	(0.10)	(0.12)	(0.11)	(0.11)	(0.07)	(0.01)
Brand 4	0.3	-0.02	0.28	0.37	0.37	0.37	0.26	0.30
		(0.38)	(0.10)	(0.13)	(0.13)	(0.13)	(0.08)	(0.02)
Preferences: Heterogeneity (Standard Dev.)								
Brand 1	0.1							0.13
								(0.13)
Brand 2	0.1							0.11
								(0.06)
Brand 3	0.1							0.10
								(0.02)
Brand 4	0.1							0.09
								(0.02)
Search Cost (Exponential Transform)								
Mean	-3	-2.89	-2.99	-2.60	-2.60	-0.260	-3.01	-3.00
		(0.65)	(0.09)	(0.09)	(0.09)	(0.09)	(0.07)	(0.03)
Standard Dev.	0.1							0.18
								(0.38)
# Observations		1,000	1,000	1,000	1,000	1,000	1,000	1,000
# Error Draws		100	1,000,000	100	100	100	100	100
# Preference Draws		n/a	n/a	n/a	n/a	n/a	n/a	100
Run Time (Seconds)		85	>576,000	86	116	1,449	450	1,255

Table B1: **Monte Carlo Simulation Results.**